

**STIC-Biotech/ChemLib**

31136

**From:** Hutzell, Paula  
**Sent:** Tuesday, December 12, 2000 5:03 PM  
**To:** STIC-Biotech/ChemLib; Bui, Phuong  
**Subject:** FW: 09/301906 sequence search

P

approved

-----Original Message-----

**From:** Bui, Phuong  
**Sent:** Tuesday, December 12, 2000 3:15 PM  
**To:** Hutzell, Paula  
**Subject:** 09/301906 sequence search

FD 4/29/99  
PTD 4/29/98

Paula,

Please forward the following search to **Barb O'Bryen** as a rush. Thanks. P.

Applic. No. 09/301906  
Phuong Bui  
AU 1638  
CM1, 9A09  
305-1996

Please search SEQ ID NO. 4. Please include interference search.

Thank you! Phuong.

**SEARCH REQUEST FORM**  
Scientific and Technical Information Center

3 11 36

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
 Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): **PAPER DISK E-MAIL**

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

**STAFF USE ONLY**

Searcher: 2013  
 Searcher Phone #: \_\_\_\_\_  
 Searcher Location: \_\_\_\_\_  
 Date Searcher Picked Up: 12-18-00  
 Date Completed: 12-18-00  
 Searcher Prep & Review Time: \_\_\_\_\_  
 Clerical Prep Time: \_\_\_\_\_  
 Online Time: \_\_\_\_\_

**Type of Search**

NA Sequence (#) \_\_\_\_\_  
 AA Sequence (#) \_\_\_\_\_  
 Structure (#) \_\_\_\_\_  
 Bibliographic \_\_\_\_\_  
 Litigation \_\_\_\_\_  
 Fulltext \_\_\_\_\_  
 Patent Family \_\_\_\_\_  
 Other \_\_\_\_\_

**Vendors and cost where applicable**

STN \_\_\_\_\_  
 Dialog \_\_\_\_\_  
 Questel/Orbit \_\_\_\_\_  
 Dr.Link \_\_\_\_\_  
 Lexis/Nexis \_\_\_\_\_  
 Sequence Systems \_\_\_\_\_  
 WWW/Internet \_\_\_\_\_  
 Other (specify) \_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 1016.61 Seconds  
(without alignments)  
1546.639 Million cell updates/sec

Title: US-09-301-906-4  
Perfect score: 360  
Sequence: 1 gtcagcggtcagtcagcgc.....agcgcgcttgacgtttggc 360

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues  
Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1:\*

2: gb\_ba2:\*

3: gb\_em:\*

4: gb\_ov:\*

5: gb\_pat:\*

6: gb\_ph:\*

7: gb\_pl1:\*

8: gb\_pl2:\*

9: gb\_pl1:\*

10: gb\_pr2:\*

11: gb\_pr3:\*

12: em\_fun:\*

13: em\_hum1:\*

14: em\_hum2:\*

15: em\_in:\*

16: em\_om:\*

17: em\_or:\*

18: em\_ov:\*

19: em\_pat:\*

20: em\_ph:\*

21: em\_pl:\*

22: em\_ro:\*

23: em\_sts:\*

24: em\_sy:\*

25: em\_un:\*

26: em\_vi:\*

27: gb\_htg1:\*

28: gb\_htg2:\*

29: gb\_in1:\*

30: gb\_in2:\*

31: em\_ba1:\*

32: em\_ba2:\*

33: em\_hum3:\*

34: em\_hum4:\*

35: gb\_pr4:\*

36: gb\_htg3:\*

37: gb\_htg4:\*

38: gb\_htg5:\*

39: gb\_htg6:\*

40: gb\_htg7:\*

41: em\_htg1:\*

42: em\_htg2:\*

43: em\_htg3:\*

44: em\_hum5:\*

45: gb\_pl3:\*

46: gb\_pr5:\*

47: gb\_htg8:\*

48: gb\_htg9:\*

49: gb\_htg10:\*

50: gb\_htg11:\*

51: gb\_htg12:\*

52: gb\_htg13:\*

53: gb\_htg14:\*

54: gb\_in3:\*

55: gb\_htg15:\*

56: gb\_htg16:\*

57: gb\_htg17:\*

58: em\_htg4:\*

59: em\_htg5:\*

60: em\_htg6:\*

61: em\_htg7:\*

62: em\_hum6:\*

63: gb\_htg18:\*

64: gb\_htg19:\*

65: gb\_ba3:\*

66: em\_htg8:\*

67: em\_htg9:\*

68: em\_htg10:\*

69: em\_htg11:\*

70: em\_htg12:\*

71: em\_htg13:\*

72: em\_htg14:\*

73: em\_htg15:\*

74: em\_htg16:\*

75: em\_htg17:\*

76: em\_htg18:\*

77: em\_htg19:\*

78: em\_htg20:\*

79: em\_htg21:\*

80: em\_htg22:\*

81: em\_htg23:\*

82: gb\_pr6:\*

83: gb\_pr7:\*

84: gb\_htg20:\*

85: gb\_htg21:\*

86: gb\_htg22:\*

87: gb\_htg23:\*

88: gb\_ro:\*

89: gb\_sts1:\*

90: gb\_sts2:\*

91: gb\_sy:\*

92: gb\_un:\*

93: gb\_vil:\*

94: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	39	10.8	3782	54	DROELAVG	M61748 Drosophila
2	38	10.6	1287	45	OSPER	X66125 O.sativa mr
3	38	10.6	1306	7	AF014467	AF014467 Oryza sat
C	38	10.6	1611	54	TBRP11	Z54340 T.brucei ge
5	36.8	10.2	677	3	AF219257	AF219257 Bos tauru
6	35.6	9.9	2140	4	XELMAPKK	D13700 Xenopus lae
7	35.2	9.8	384	11	AF219258	AF219258 Homo sapi
C	35	9.7	160013	11	AL133325	AL133325 Human DNA
C	34.2	9.5	753	88	MMU18673	U18673 Mus musculu
C	33.8	9.4	9003	38	AC014530	AC014530 Drosophil
C	33.8	9.4	83188	37	AC013383	AC013383 Homo sapi
C	33.8	9.4	143816	38	AC013598	AC013598 Homo sapi













J. Virol. 69 (4), 2501-2507 (1995)

95191029

MEDLINE

2 (bases 1 to 753)

AUTHORS

Shackelford, G.M.

TITLE

Direct Submission

JOURNAL

Submitted (15-DEC-1994) Gregory M. Shackelford, Division

Hematology-Oncology, Univ. of Southern California and Childrens

Hospital Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA

90027-6016, USA

FEATURES

source

1..753

/organism="Mus musculus"

/db\_xref="taxon:10090"

/sex="female"

/tissue\_type="mammary tumor"

85..699

/gene="Pgf-8"

85..699

/note="alternatively spliced mRNA; isoform Pgf-8a"

/codon\_start=1

/product="fibroblast growth factor-8a"

/protein\_id="AAA65387.1"

/db\_xref="GI:619920"

/translation="MGSPRSALSCLLHLLVLCQAQHVREOSLVTDLSRLRLRTQY

LYRTSGKHVOVLANKRINAMAEQDPFAKLIVETDTFGSRVVRGAETGLYICNKK

GKLIASNKCKDCQVETIVLENNYATLONAKYEGWYMAFTKGRPRGSKTRQHRE

VHFKRLPRGHHTQSLRFEFLNPPFTRSLRGSQRTWAPEPR"

BASE COUNT 152 a 265 c 227 g 109 t

ORIGIN

Query Match 9.5%; Score 34.2; DB 88; Length 753;

Best Local Similarity 48.2%; Pred. No. 21;

Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 148 cgggtagtcgcgtatgcccccgcagcgagcagcctaaacccgctgttaagcaagat 207

Db 207 CCGGATGAGCGCGCGCTGAGCTGATGTCACACAGGCTCTGCTCCCTCACATGCTGGGC 148

Qy 208 gaggcagaagcccaagcgaagcgttcacattggctgttaagcacaagcgttggtc 267

Db 147 TTGGAGGAGAGAGACGACGAGTGAACACAGCAGGAGCTCAGCGGGGGGTGCC 88

Qy 268 cagctaccattcctaaaaaacaggaagcactggagcagcgaatcagtcaccacaag 327

Db 87 CAGCGCGCGCGCGCGGAGGAGCTGAGAGCCGAGCGGGGTACGCGCTCCACCTGGAG 28

Qy 328 tcgttgaggaggaagggccg 346

Db 27 GCCGCGGGGGACAGCCG 9

RESULT 10

AC014530/c

LOCUS

AC014530 9003 bp DNA HTG 16-NOV-1999

DEFINITION

Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered

pieces.

ACCESSION

AC014530

VERSION

AC014530.1 GI:6436805

KEYWORDS

HTG; HTGS\_PHASE2.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 9003)

AUTHORS

Adams, M. and Venter, J.C.

TITLE

Direct Submission

JOURNAL

Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

COMMENT

This sequence was identified as CDM:10212184 by the submitter.

For further information on this sequence e-mail to fly@celera.com.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

FEATURES

source

1..9003

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

BASE COUNT 2485 a 1945 c 1940 g 2633 t

ORIGIN

Query Match 9.4%; Score 33.8; DB 38; Length 9003;

Best Local Similarity 49.7%; Pred. No. 30;

Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 170 tcaagcgcagcagcctaaacccgctgttaagcagaatgagcagaagcccaagcgcaag 229

Db 8682 TCAAGCGTCTGCAGCAGCAGCGCCATGCAACACGACGACGACGACGACCATCAAC 8623

Qy 230 cgtcacattggcgtttaagcacaagcgtttggcgtccagctaccacttctctaaaaac 289

Db 8622 AACACGACGACGACGACGACCAAAATGGTGTGTTCCCAATTTCAACAAACACAGC 8563

Qy 290 aggaagcactggagcagcgaatcagtcacacacagcgtgttgagagagaag 342

Db 8562 AGCAGCAACAGGTGGCCGACGACGACGACGACGACGACGACGACGACGACG 8510

RESULT 11

AC013383

LOCUS

AC013383 83188 bp DNA HTG 13-JUL-2000

DEFINITION

Homo sapiens chromosome 2 clone RP11-396G4 map 2, LOW-PASS SEQUENCE

SAMPLING.

ACCESSION

AC013383

VERSION

AC013383.2 GI:9123916

KEYWORDS

HTG; HTGS\_PHASE0.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 83188)

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens chromosome 2, clone RP11-396G4

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 83188)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galan, J., Gardyna, S., Grant, G., Hago, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6289163.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3790  
Center clone name: 396\_G\_4

-----  
\* NOTE: This record contains 87 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 635: contig of 635 bp in length  
\* 636 735: gap of 100 bp  
\* 736 1579: contig of 844 bp in length  
\* 1580 1679: gap of 100 bp  
\* 1680 2558: contig of 879 bp in length  
\* 2559 2658: gap of 100 bp  
\* 2659 3511: contig of 853 bp in length  
\* 3512 3611: gap of 100 bp  
\* 3612 4461: contig of 850 bp in length  
\* 4462 4561: gap of 100 bp  
\* 4562 5399: contig of 838 bp in length  
\* 5400 5499: gap of 100 bp  
\* 5500 6399: contig of 900 bp in length  
\* 6400 6499: gap of 100 bp  
\* 6500 7355: contig of 856 bp in length  
\* 7356 7455: gap of 100 bp  
\* 7456 8268: contig of 813 bp in length  
\* 8269 8368: gap of 100 bp  
\* 8369 9239: contig of 871 bp in length  
\* 9240 9339: gap of 100 bp  
\* 9340 10190: contig of 851 bp in length  
\* 10191 10290: gap of 100 bp  
\* 10291 11147: contig of 857 bp in length  
\* 11148 11247: gap of 100 bp  
\* 11248 12131: contig of 884 bp in length  
\* 12132 12231: gap of 100 bp  
\* 12232 13105: contig of 874 bp in length  
\* 13106 13205: gap of 100 bp  
\* 13206 14071: contig of 866 bp in length  
\* 14072 14171: gap of 100 bp  
\* 14172 15049: contig of 878 bp in length  
\* 15050 15149: gap of 100 bp  
\* 15150 16013: contig of 864 bp in length  
\* 16014 16113: gap of 100 bp  
\* 16114 16981: contig of 868 bp in length  
\* 16982 17081: gap of 100 bp  
\* 17082 17942: contig of 861 bp in length  
\* 17943 18042: gap of 100 bp  
\* 18043 18903: contig of 861 bp in length  
\* 18904 19003: gap of 100 bp  
\* 19004 19879: contig of 876 bp in length  
\* 19880 19979: gap of 100 bp  
\* 19980 20834: contig of 855 bp in length  
\* 20835 20934: gap of 100 bp  
\* 20935 21787: contig of 853 bp in length  
\* 21788 21887: gap of 100 bp  
\* 21888 22742: contig of 855 bp in length  
\* 22743 22842: gap of 100 bp  
\* 22843 23691: contig of 849 bp in length  
\* 23692 23791: gap of 100 bp  
\* 23792 24657: contig of 866 bp in length  
\* 24658 24757: gap of 100 bp  
\* 24758 25630: contig of 873 bp in length  
\* 25631 25730: gap of 100 bp  
\* 25731 26584: contig of 854 bp in length  
\* 26585 26684: gap of 100 bp  
\* 26685 27579: contig of 895 bp in length  
\* 27580 27679: gap of 100 bp  
\* 27680 28553: contig of 874 bp in length

28554 28653: gap of 100 bp  
28654 29505: contig of 852 bp in length  
29506 29605: gap of 100 bp  
29606 30433: contig of 828 bp in length  
30434 30533: gap of 100 bp  
30534 31389: contig of 856 bp in length  
31390 31489: gap of 100 bp  
31490 32354: contig of 865 bp in length  
32355 32454: gap of 100 bp  
32455 33314: contig of 860 bp in length  
33315 33414: gap of 100 bp  
33415 34265: contig of 851 bp in length  
34266 34365: gap of 100 bp  
34366 35234: contig of 869 bp in length  
35235 35334: gap of 100 bp  
35335 36206: contig of 872 bp in length  
36207 36306: gap of 100 bp  
36307 37170: contig of 864 bp in length  
37171 37270: gap of 100 bp  
37271 38143: contig of 873 bp in length  
38144 38243: gap of 100 bp  
38244 39107: contig of 864 bp in length  
39108 39207: gap of 100 bp  
39208 40060: contig of 853 bp in length  
40061 40160: gap of 100 bp  
40161 41019: contig of 859 bp in length  
41020 41119: gap of 100 bp  
41120 41975: contig of 856 bp in length  
41976 42075: gap of 100 bp  
42076 42931: contig of 856 bp in length  
42932 43031: gap of 100 bp  
43032 43864: contig of 833 bp in length  
43865 43964: gap of 100 bp  
43965 44825: contig of 861 bp in length  
44826 44925: gap of 100 bp  
44926 45813: contig of 888 bp in length  
45814 45913: gap of 100 bp  
45914 46779: contig of 866 bp in length  
46780 46879: gap of 100 bp  
46880 47715: contig of 836 bp in length  
47716 47815: gap of 100 bp  
47816 48694: contig of 879 bp in length  
48695 48794: gap of 100 bp  
48795 49621: contig of 827 bp in length  
49622 49721: gap of 100 bp  
49722 50598: contig of 877 bp in length  
50599 50698: gap of 100 bp  
50699 51562: contig of 864 bp in length  
51563 51662: gap of 100 bp  
51663 52540: contig of 878 bp in length  
52541 52640: gap of 100 bp  
52641 53513: contig of 873 bp in length  
53514 53613: gap of 100 bp  
53614 54474: contig of 861 bp in length  
54475 54574: gap of 100 bp  
54575 55434: contig of 860 bp in length  
55435 55534: gap of 100 bp  
55535 56424: contig of 890 bp in length  
56425 56524: gap of 100 bp  
56525 57450: contig of 926 bp in length  
57451 57550: gap of 100 bp  
57551 58410: contig of 860 bp in length  
58411 58510: gap of 100 bp  
58511 59329: contig of 819 bp in length  
59330 59429: gap of 100 bp  
59430 60276: contig of 847 bp in length  
60277 60376: gap of 100 bp  
60377 61245: contig of 869 bp in length  
61246 61345: gap of 100 bp  
61346 62198: contig of 853 bp in length  
62199 62298: gap of 100 bp  
62299 63171: contig of 873 bp in length  
63172 63271: gap of 100 bp

\* 63272 64148: contig of 877 bp in length
\* 64149 64248: gap of 100 bp
\* 64249 65108: contig of 860 bp in length
\* 65109 65208: gap of 100 bp
\* 65209 66052: contig of 844 bp in length
\* 66053 66152: gap of 100 bp
\* 66153 66983: contig of 831 bp in length
\* 66984 67083: gap of 100 bp
\* 67084 67948: contig of 865 bp in length
\* 67949 68048: gap of 100 bp
\* 68049 68929: contig of 881 bp in length
\* 68930 69029: gap of 100 bp

Query Match 9.4%; Score 33.8; DB 37; Length 83188;
Best Local Similarity 50.3%; Pred. No. 32;
Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 114 ttttttaagtacttctgtacactaccacggtagtgcgtatgccccctgcaa 173
DB 55907 TGTCTTTAGCCATTATCTGTAATAATACACAGAGTCTCTTTGGTTTGTAGTGA AAA 55966

QY 174 ggcgcagcgcctaaacggcgttaagcaagatgacgagaccccaaacgcaagcgtc 233
DB 55967 CAATGAGGAACACACACCTTAAAGATGTAACGAGAGCAACAAACGGAATGTC 56026

QY 234 acattggcgtttaagcacaacagctgttgccgtccacgtaccact 278
DB 56027 TCTCTTTGCAGACACACCTTCATCTGTACACCTGCACATAGCCCT 56071

RESULT 12
AC013598/c
LOCUS AC013598 143816 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-9L19, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC013598
VERSION 1 (bases 1 to 143816)
KEYWORDS HTG; HTCS\_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143816)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelhiano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lhoczyk,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melorim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6715989.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence\_submissions@genome.wi.mit.edu
----- Project Information

Center project name: L3017
Center clone name: 9\_L\_19
-----
\* NOTE: This record contains 146 individual
\* sequencing reads that have not been assembled into
\* contigs. Runs of N are used to separate the reads
\* and the order in which they appear is completely
\* arbitrary. Low-pass sequence sampling is useful for
\* identifying clones that may be gene-rich and allows
\* overlap relationships among clones to be deduced.
\* However, it should not be assumed that this clone
\* will be sequenced to completion. In the event that
\* the record is updated, the accession number will
\* be preserved.
\* 1 839: contig of 839 bp in length
\* 840 939: gap of 100 bp
\* 940 1832: contig of 893 bp in length
\* 1833 1932: gap of 100 bp
\* 1933 2819: contig of 887 bp in length
\* 2820 2919: gap of 100 bp
\* 2920 3866: contig of 947 bp in length
\* 3867 3966: gap of 100 bp
\* 3967 4853: contig of 887 bp in length
\* 4854 4953: gap of 100 bp
\* 4954 5801: contig of 848 bp in length
\* 5802 5901: gap of 100 bp
\* 5902 6767: contig of 866 bp in length
\* 6768 6867: gap of 100 bp
\* 6868 7734: contig of 867 bp in length
\* 7735 7834: gap of 100 bp
\* 7835 8677: contig of 843 bp in length
\* 8678 8777: gap of 100 bp
\* 8778 9633: contig of 856 bp in length
\* 9634 9733: gap of 100 bp
\* 9734 10597: contig of 864 bp in length
\* 10598 10697: gap of 100 bp
\* 10698 11585: contig of 888 bp in length
\* 11586 11685: gap of 100 bp
\* 11686 12556: contig of 871 bp in length
\* 12557 12656: gap of 100 bp
\* 12657 13519: contig of 863 bp in length
\* 13520 13619: gap of 100 bp
\* 13620 14493: contig of 874 bp in length
\* 14494 14593: gap of 100 bp
\* 14594 15459: contig of 866 bp in length
\* 15460 15559: gap of 100 bp
\* 15560 16393: contig of 834 bp in length
\* 16394 16493: gap of 100 bp
\* 16494 17340: contig of 847 bp in length
\* 17341 17440: gap of 100 bp
\* 17441 18288: contig of 848 bp in length
\* 18289 18388: gap of 100 bp
\* 18389 19230: contig of 842 bp in length
\* 19231 19330: gap of 100 bp
\* 19331 20202: contig of 872 bp in length
\* 20203 20302: gap of 100 bp
\* 20303 21180: contig of 878 bp in length
\* 21181 21280: gap of 100 bp
\* 21281 22160: contig of 880 bp in length
\* 22161 22260: gap of 100 bp
\* 22261 23159: contig of 899 bp in length
\* 23160 23259: gap of 100 bp
\* 23260 24140: contig of 881 bp in length
\* 24141 24240: gap of 100 bp
\* 24241 25109: contig of 869 bp in length
\* 25110 25209: gap of 100 bp
\* 25210 26053: contig of 844 bp in length
\* 26054 26153: gap of 100 bp
\* 26154 27016: contig of 863 bp in length
\* 27017 27116: gap of 100 bp
\* 27117 28063: contig of 947 bp in length
\* 28064 28163: gap of 100 bp
\* 28164 29021: contig of 858 bp in length



Quality coverage: 4.66 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3337: contig of 3337 bp in length  
\* 3338 3437: gap of unknown length  
\* 3438 10367: contig of 6930 bp in length  
\* 10368 10468: gap of unknown length  
\* 10468 22403: contig of 11936 bp in length  
\* 22404 22503: gap of unknown length  
\* 22504 31981: contig of 9478 bp in length  
\* 31982 32081: gap of unknown length  
\* 32082 44508: contig of 12427 bp in length  
\* 44509 44608: gap of unknown length  
\* 44609 71086: contig of 26478 bp in length  
\* 71087 71186: gap of unknown length  
\* 71187 99473: contig of 28287 bp in length  
\* 99474 99573: gap of unknown length  
\* 99574 134831: contig of 35258 bp in length  
\* 134832 134931: gap of unknown length  
\* 134932 174101: contig of 39170 bp in length.

## FEATURES

## source

1. 174101  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-491M19"

## misc\_feature

1. 3337  
/note="assembly\_name:Contig21"  
clone\_end:SP6  
vector\_side:right"  
3438..10367  
/note="assembly\_name:Contig22"  
10468..22403  
/note="assembly\_name:Contig23"  
22504..31981  
/note="assembly\_name:Contig24"  
32082..44508  
/note="assembly\_name:Contig25"  
44609..71086  
/note="assembly\_name:Contig26"  
71187..99473  
/note="assembly\_name:Contig27"  
99574..134831  
/note="assembly\_name:Contig28"  
134932..174101  
/note="assembly\_name:Contig29"

BASE COUNT 51725 a 35565 c 34703 g 51300 t 808 others  
ORIGIN

Query Match 9.4%; Score 33.8; DB 36; Length 174101;  
Best Local Similarity 50.3%; Pred. No. 33;  
Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
Qy 114 tttcttaagtacttctgtctacactaccacgggtagtcgtatgccccctcaa 173  
Db 82749 TCTTTTAGCCATTCCTGTAATAATACACAGAGTCTCTTTGTTAGTGA 82808  
Qy 174 gcgcagagccttaacggcgtttaagcaagatgagcagaagcccaaacggaagcgc 233  
Db 82809 CAATGAGGAACACACACCTTAAAGAAATGTAACGAGAGCAACAAACGGGAATGTC 82868  
Qy 234 acatggggcgttaagcccaacagcgtgttgccgtccacgtaccact 278  
Db 82869 TCTCTTTGCCAGACACACCTTCATCTGTCATACCTGCACATAGCCCT 82913

## RESULT 14

AC018877  
LOCUS  
DEFINITION  
AC018877  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC018877 195347 bp DNA HTG  
Homo sapiens chromosome 2 clone RP11-295K1, WORKING DRAFT SEQUENCE,  
27 unordered pieces.  
AC018877 GI:9719807  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 195347)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 195347)  
Waterston,R.H.  
Direct Submission  
Submitted (21-DEC-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Aug 7, 2000 this sequence version replaced gi:6624094.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: HNH0295K01  
----- Summary Statistics -----  
Sequencing vector: M13; 63%  
Sequencing vector: plasmid; 37%  
Chemistry: Dye-primer ET; 63% of reads  
Chemistry: Dye-terminator Big Dye; 37% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 176143 bases at least Q40  
Consensus quality: 182219 bases at least Q30  
Consensus quality: 185619 bases at least Q20  
Insert size: 194000; agarose-fp  
Insert size: 192747; sum-of-contigs  
Quality coverage: 3.36 in Q20 bases; agarose-fp  
Quality coverage: 3.45 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1250: contig of 1250 bp in length  
\* 1251 1350: gap of unknown length  
\* 1351 2427: contig of 1077 bp in length  
\* 2428 2527: gap of unknown length  
\* 2528 4162: contig of 1635 bp in length  
\* 4163 4262: gap of unknown length  
\* 4263 6032: contig of 1770 bp in length  
\* 6033 6132: gap of unknown length  
\* 6133 9005: contig of 2873 bp in length  
\* 9006 11520: contig of 2415 bp in length  
\* 11521 11620: gap of unknown length  
\* 11621 13993: contig of 2373 bp in length  
\* 13994 14093: gap of unknown length  
\* 14094 16557: contig of 2464 bp in length  
\* 16558 16658: gap of unknown length  
\* 16659 19112: contig of 2455 bp in length  
\* 19113 19212: gap of unknown length  
\* 19213 21091: contig of 1879 bp in length  
\* 21092 21191: gap of unknown length



## COMMENT

Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.

For a graphical representation of this sequence and its analysis see:-

<http://webace.sanger.ac.uk/cgi-bin/display?db=wormaceclass-sequence;object=y40b1b>  
 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.  
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

IMPORTANT: This sequence is not the entire insert of clone Y40B1B. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true right end of clone W05H12 is at 102 in this sequence. The true right end of clone Y40B1 is at 29313 in this sequence. The start of this sequence (1..102) overlaps with the end of sequence Z82072.

The end of this sequence (29204..29313) overlaps with the start of sequence AL132847.

## FEATURES

```

source
1. 29313
   Location/Qualifiers
   /organism="Caenorhabditis elegans"
   /db_xref="taxon:6239"
   /chromosome="I"
   /clone="Y40B1B"
   complement(join(2625..3579,3640..3716))
   /gene="Y40B1B.3"
   complement(join(2625..3579,3640..3716))
   /gene="Y40B1B.3"
   /codon_start=1
   /protein_id="CAA21605.1"
   /db_xref="GI:3880909"
   /db_xref="SPTREMBL:Q9XWP5"
   /translation="MSTFPIILLPQKSLQNVLRIMQPLELYFVSVKTKAHTVDLN
   LKASRLVLKNGECAMIEAVFKDGNVTGVEISTEACDCSEVMYPLVGYEDGSY
   DSSHADFIVKTLNLAQVIAQVIFHKPKISQLYHENVYICDQMLEIKILDFEKIE
   FWITFDYFLIFNGYREVDFQIHLRLVRLVILCRFVDSHLPNVLASNGT
   QFKCSTVPKISINMELRHWTGNSRPLEICIRISIEAPESEESTYEEVLKGTHYO
   TEPTDSVKTEPILNYHNNLCFNEHLKQTGTGIVRCGYEKKANVVKLPETPFESE
   VLFQMSLIIV"
   complement(join(5128..5217,6728..6929,6979..7058))
   /gene="Y40B1B.7"
   complement(join(5128..5217,6728..6929,6979..7058))
   /gene="Y40B1B.7"
   /note="cDNA EST yk546c3.3 comes from this gene
   cDNA EST yk372d6.3 comes from this gene
   cDNA EST yk372d6.5 comes from this gene
   cDNA EST yk546c3.5 comes from this gene
   cDNA EST yk562g2.5 comes from this gene"
   /codon_start=1
   /protein_id="CAA21606.1"
   /db_xref="GI:3880910"
   /db_xref="SPTREMBL:Q9XWP4"
   /translation="MSTGANLLVMNDCKSNRWKTKQEKHSEIKVKYTLKSTWDKK
   MELKAKDMVKRVODNIREKQVQERQEKKEKRLNERKAEIVQKTKIKHL
   KTKYKRLRSLQMRDTQVTK"
   join(7514..7595,7679..7823,9391..9684,9733..9790,
   9847..9995,10578..10908,12393..12628,12679..12798,
   12885..12972)
   /gene="Y40B1B.8"
   join(7514..7595,7679..7823,9391..9684,9733..9790,
   9847..9995,10578..10908,12393..12628,12679..12798,
   12885..12972)
   /gene="Y40B1B.8"
   /note="cDNA EST yk597b12.3 comes from this gene
   cDNA EST yk235e7.3 comes from this gene"

```

```

CDNA EST yk235e7.5 comes from this gene
CDNA EST yk382c6.5 comes from this gene
CDNA EST yk597b12.5 comes from this gene
/codon_start=1
/protein_id="CAA21607.2"
/db_xref="GI:5824762"
/db_xref="SPTREMBL:Q9U2J0"
/translation="MPTQFIRNRSAFPOQTNNDDIDFQYDNFSSSHQLPSTAGYS
EHSNQFNQNSHPPEKATDPLAGALLGSDVITKSLIHPGCVLRQCQVHQFAGS
LHLTVTLIPVCNSVAKOGIOTFWKGAIGSSVLMGLTNVTETVLGMLGRLTFVYN
GSAEKYKHIVLKGVTEFTMTPTFYISSTETVRSSEIGGEDNKILDVLVKGVDRMY
PLTSGRDSSRRSIIHLAIPTAGFQWGHYMLQFALYHOFERMAKRYVNSKPAEKTY
HDLFQMAQTSMTMLDLIPFETIVHRMTIQGTRTLLDNMDTGLSAYSMVKISG
YDCVQVLETFEGWALYAGVGLLEYSLHQGLQQLVRACFRGSELLKKAQGHVP
VQSSQITPPIISAKSFSNPSINSFSAAGALLSPMTPRKHSHTPPKDPPTFTGETV
SQIGSPYQMPDRPNVFGSPPNRTRATHDTSQSLPQLNIRPSSGGPPFSG"
join(14099..14519,15347..15429,16169..16303)
/gene="Y40B1B.5"
join(14099..14519,15347..15429,16169..16303)
/gene="Y40B1B.5"
/note="predicted using Genefinder
CDNA EST yk4e5.5 comes from this gene
CDNA EST yk464d3.5 comes from this gene
CDNA EST yk414e6.5 comes from this gene
CDNA EST yk490e10.5 comes from this gene
CDNA EST yk90f3.5 comes from this gene
CDNA EST yk335g9.5 comes from this gene
CDNA EST yk325b12.5 comes from this gene
CDNA EST yk589e10.5 comes from this gene
CDNA EST yk590a9.5 comes from this gene
CDNA EST yk625g8.5 comes from this gene
CDNA EST yk625g9.5 comes from this gene
CDNA EST yk636a11.5 comes from this gene
CDNA EST yk673f12.5 comes from this gene
CDNA EST yk677d6.5 comes from this gene"
/codon_start=1
/protein_id="CAA21603.1"
/db_xref="GI:3880907"
/db_xref="SPTREMBL:Q9XWP7"
/translation="MSDWEDDDFEPEVSTFKRAEPAPKVIKDAPPPKAPAAAPK
TKUAAPTFAMESLGRLETSAREAIQKNDLALRDLFGDDSAEAIKMTVANIHS
KADFYRWGRVGFSLASCSCYGDMLKLLTSVDEMTPAELIKMTVTLQIQISA
KTAETKATKATTAAPAAANKNAKATLVTKGNSDMYDDYGDQDDQDDYD"
join(16662..16740,17501..17730,19395..19700,20983..21625,
22437..22635,23692..23868,25405..25782,28224..28524)
/gene="Y40B1B.6"
join(16662..16740,17501..17730,19395..19700,20983..21625,
22437..22635,23692..23868,25405..25782,28224..28524)
/gene="Y40B1B.6"
/note="predicted using Genefinder
CDNA EST yk122d5.5 comes from this gene
CDNA EST yk141d9.5 comes from this gene
CDNA EST yk290b12.5 comes from this gene
CDNA EST yk251c1.5 comes from this gene
CDNA EST yk174e9.5 comes from this gene
CDNA EST yk669h5.3 comes from this gene
CDNA EST yk379a6.5 comes from this gene
CDNA EST yk225h10.5 comes from this gene
CDNA EST yk533d5.5 comes from this gene
CDNA EST yk584a9.5 comes from this gene
CDNA EST yk606g3.5 comes from this gene"
/codon_start=1
/protein_id="CAA21604.1"
/db_xref="GI:3880908"
/db_xref="SPTREMBL:Q9XWP6"
/translation="MSSDTGSEYLDDEIRGDELPSIDNALAAASAARLPDRPTD
HELAPPELWEHKTAVEYVLLRNSTLATWYNPLKECTALDVRNNVFPFNSDLDLI
QNAVYLSRHGLINFGYVYRSTKISRLFVDRRSVIVIGAGAAGISAAITOLESPGDV
IYVLRNCGIGRIHSPKSKSGEIMETGGDTLRLKIESPMTALLHOVNEFHGVDFDTS
VVEGRPLNEEKIHLFDHYKSAHGALNYOAHOCERHDDOGSFSISROQVLENLSMCE
RGTLLIKYFNCKSLFVARAREHHFNQKRLTALMAENQLKMEEGLEQDDPVLRL
RSLKRDIAISLEKFEFADAFETADNHWORLNHPQAKOYMHFSGSEFATFNFMLGPEE
YLVGAQLEKVFQSCDSMQNKENGVAARLTGEGIAELLTQLSEKKRLDIRLKHRLDIDY

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 25.1 Seconds  
(without alignments)  
2169.194 Million cell updates/sec

Title: US-09-301-906-4  
Perfect score: 360  
Sequence: 1 gtcagcggctcagtcagcgc.....aggccgccttgacgttggtgc 360

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620496 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*

7: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	29.4	8.2	11748	2	US-08-611-107-30
C 2	29	8.1	1560	6	PCT-US94-09789-1
C 3	29	8.1	2158	2	US-08-698-551-1
C 4	29	8.1	2158	3	US-08-602-228-1
C 5	29	8.1	2158	3	US-08-649-341A-1
C 6	29	8.1	2158	3	US-08-494-440B-1
C 7	29	8.1	2158	3	US-08-533-901B-1
C 8	29	8.1	2158	3	US-08-839-032A-1
C 9	29	8.1	2158	4	US-08-839-031A-1
C 10	29	8.1	2158	6	PCT-US95-12724-1
C 11	29	8.1	2735	3	US-08-698-551-11
C 12	29	8.1	2735	3	US-08-602-228-11
C 13	29	8.1	2735	3	US-08-494-440B-11
C 14	29	8.1	2735	3	US-08-533-901B-11
C 15	29	8.1	2735	3	US-08-839-032A-11
C 16	29	8.1	2735	4	US-08-839-031A-11
C 17	29	8.1	2735	6	PCT-US95-12724-11
C 18	29	8.1	6002	2	US-08-698-551-15
C 19	29	8.1	6002	3	US-08-602-228-15
C 20	29	8.1	6002	3	US-08-839-032A-15
C 21	29	8.1	7360	2	US-08-611-107-9
C 22	29	8.1	7360	4	US-08-422-560A-9
C 23	28.2	7.8	545	1	US-07-915-966C-1
C 24	28.2	7.8	545	4	US-08-771-182-1
C 25	28.2	7.8	545	5	US-08-853-194-1
C 26	28	7.8	435	2	US-08-694-579-2

C	27	28	7.8	435	3	US-08-948-155-2	Sequence 2, Appli
	28	28	7.8	1689	5	US-08-335-844A-6	Sequence 6, Appli
	29	28	7.8	3084	5	US-08-335-844A-19	Sequence 19, Appli
	30	27.6	7.7	2012	3	US-08-484-200-3	Sequence 3, Appli
	31	27.2	7.6	2223	1	US-08-317-522A-8	Sequence 8, Appli
	32	27.2	7.6	2223	1	US-08-439-818A-8	Sequence 8, Appli
	33	27.2	7.6	2223	3	US-08-751-965-8	Sequence 8, Appli
	34	27.2	7.6	2223	3	US-08-738-975-8	Sequence 8, Appli
	35	27.2	7.6	2223	4	US-08-728-626-8	Sequence 8, Appli
	36	27.2	7.6	2223	5	US-08-808-599A-8	Sequence 8, Appli
	37	26.8	7.4	1524	7	5512669-1	Patent No. 5512669
	38	26.8	7.4	1525	7	529279-1	Patent No. 529279
	39	26.6	7.4	1045	5	US-08-859-167-3	Sequence 3, Appli
	40	26.6	7.4	1045	5	US-09-109-273-3	Sequence 3, Appli
C	41	26.4	7.3	1604	2	US-08-694-579-1	Sequence 1, Appli
C	42	26.4	7.3	1604	3	US-08-948-155-1	Sequence 1, Appli
	43	26.2	7.3	1031	1	US-08-181-271A-106	Sequence 106, App
	44	26.2	7.3	1031	1	US-08-449-315-106	Sequence 106, App
	45	26.2	7.3	1031	1	US-08-444-803-106	Sequence 106, App

ALIGNMENTS

RESULT 1

US-08-611-107-30/c

; Sequence 30. Application US/08611107

; Patent No. 5801233

; GENERAL INFORMATION:

; APPLICANT: Haselkorn, Robert

; APPLICANT: Gornicki, Piotr

; TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING

; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE AND USES

; TITLE OF INVENTION: THEREFOR

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: United States of America

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/611,107

; FILING DATE: Concurrently Herewith

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US SN 07/956,700

; FILING DATE: 02-OCT-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/09340

; FILING DATE: 30-SEP-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US SN 08/422,560

; FILING DATE: 14-APR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Kitchell, Barbara S.

; REGISTRATION NUMBER: 33,928

; REFERENCE/DOCKET NUMBER: ARCD:221

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (512) 474-7577

; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 11748 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

Db 141 CAGGAAAGACAAAGAGGGTGTCTCTATGTAGGCTCAAAACCAAGGAGGGAGTGGTGC 200

APPLICANT: Lin, Lih-Ling  
APPLICANT: Chen, Jennifer H.  
APPLICANT: Schievella, Andrea  
APPLICANT: Graham, James  
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND  
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA

```

/
/
/      COMPUTER READABLE FORM:
/      MEDIUM TYPE: Floppy disk
/      COMPUTER: IBM PC compatible
/      OPERATING SYSTEM: PC-DOS/MS-DOS
/      SOFTWARE: PatentIn Release #1.0, Version #1.25
/      CURRENT APPLICATION DATA:
/      APPLICATION NUMBER: US/08/602,228
/

```

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A,  
REGISTRATION NUMBER: 32,724  
REFERENCE/DOCKET NUMBER: G15232C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5951

```

; INFORMATION FOR SEQ ID NO: 1:
; =====
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 2158 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
;     HYPOTHETICAL: NO
;     FEATURE:
;         NAME/KEY: CDS
;         LOCATION: 2..1231
;         US-08-602-228-1

```

Query Match 8.1%; Score 29; DB 3; Length 2158;  
Best Local Similarity 53.0%; Pred. No. 1.8;  
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 195 tgттаgcaagatgagcagaagccaaacggcaagctcacattggcgctgttaagccaac 254  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 530 TGATTGGTGGGTGTCACAAGCCCAANTGCCGACTTTCCCATTAGGGCGCTCAGCTTCT 471

255 agctgttggtccacgtaccacttcctataaaaaaacaggaagcactggagccagcgca 311  
 470 TCGCGATGTCATCTTATTTACCTTCATCAGCAGCATGTAGAGATGAGGTGTGCA 414

RESULT 5  
US-08-649-341A-1/c  
; Sequence 1, Application US/08649341A  
; Patent No. 5847099  
; GENERAL INFORMATION:

```

/ GENERAL INFORMATION:
/
/ APPLICANT: Lin, Lih-Ling
/ APPLICANT: Chen, Jennifer H.
/
/ TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
/
/ TITLE OF INVENTION: PROTEINS
/
/ NUMBER OF SEQUENCES: 9
/
/ CORRESPONDENCE ADDRESS:
/

```

ADDRESS: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA

```

, ZIP: 02140
,
, COMPUTER READABLE FORM:
,
, MEDIUM TYPE: Floppy disk
,
, COMPUTER: IBM PC compatible
,
, OPERATING SYSTEM: PC-DOS/MS-DOS
,
, SOFTWARE: PatentIn Release #1.0, V
,
, CURRENT APPLICATION DATA:
,
, APPLICATION NUMBER: US/08/649,341A
,
, FILING DATE:
,
, CLASSIFICATION: 435
,
, ATTORNEY/AGENT INFORMATION:
,
, NAME: Brown, Scott A.,
,
, REGISTRATION NUMBER: 32,724
,
, REFERENCE/DOCKET NUMBER: G15232-FW
,
, TELECOMMUNICATION INFORMATION:
,
, TELEPHONE: (617) 498-8224
,
, TELEFAX: (617) 876-5851
,
, INFORMATION FOR SEQ ID NO: 1:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 2158 base pairs
,
, TYPE: nucleic acid
,
, STRANDEDNESS: double
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: cDNA
,
, HYPOTHETICAL: NO
,
, FEATURE:
,
, NAME/KEY: CDS
,
, LOCATION: 2..1231
,
, US-08-649-341A-1

```

Query Match 8.1%; Score 29; DB 3; Length 2158;  
Best Local Similarity 53.0%; Pred. No. 1.8;  
Matches 62; Conservative 0; Mismatches 55; Indels

Qy	195	tgtaaagaagatgagcagaagcccaaacggaagcgtcacatttggcgtgtaagccaac	254
Db	530	TGATTTGCTGCTGTACACAAAGCCCAATGTCGCACTTTCCCATTTAGCGGCTCACCTTCT	471
Qy	255	agctgttggcgtccacgtacacactctctaaaaacaggaaagcactggagccagcgca	311
Db	470	TGCGGATGTCATCTTTATTTACCTTTCATCAGCAGCATGTAGGAGATGAGCTTCTGCA	414

RESULT 6  
US-08-494-440B-1/c  
; Sequence 1, Application US/08494440B  
; Patent No. 5849501

APPLICANT: Lin, Lih-Ling  
APPLICANT: Chen, Jennifer H.  
APPLICANT: Schievella, Andrea  
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND  
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:

```

; ADDRESS:  Genetics Institute, Inc
; STREET:  87 CambridgePark Drive
; CITY:    Cambridge
; STATE:   Massachusetts
; COUNTRY: USA
; ZIP:     02140
;

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatenIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/494,440B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.

Query Match 8.1%; Score 29; DB 3; Length 2158;  
Best Local Similarity 53.0%; Pred. NO. 1.8;  
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

5' TCGGGATGTCATTCCTTATTTACCTTCATCAGCAGCATGTAGGAGATGAGGTTGT

RESULT 8

S-08-839-032A-1/C  
Sequence 1, Application US/08839032A  
Patent No. 5891675  
GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling  
APPLICANT: Chen, Jennifer H.  
APPLICANT: Schievella, Andrea  
APPLICANT: Graham, James

RESULT 7 .  
US-08-533-901B-1/c

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Ver  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/839, 032A

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/839,032A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWN, SCOTT A,  
PROFESSIONAL NUMBER: 33,73

REGISTRATION NUMBER: 347/24  
REFERENCE/DOCKET NUMBER: G15232DDIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2158 base pairs

LIFE: NUCLEIC ACIDS  
 STRANDEDNESS: DOUBLE  
 TOPOLOGY: LINEAR

MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO

NAME/KEY: CDS

LOCATION: 2..1231  
S-08-839-032A-1

Query Match	Best Local Similarity
0	0.00
1	0.00
2	0.00
3	0.00
4	0.00
5	0.00
6	0.00
7	0.00
8	0.00
9	0.00
10	0.00
11	0.00
12	0.00
13	0.00
14	0.00
15	0.00
16	0.00
17	0.00
18	0.00
19	0.00
20	0.00
21	0.00
22	0.00
23	0.00
24	0.00
25	0.00
26	0.00
27	0.00
28	0.00
29	0.00
30	0.00
31	0.00
32	0.00
33	0.00
34	0.00
35	0.00
36	0.00
37	0.00
38	0.00
39	0.00
40	0.00
41	0.00
42	0.00
43	0.00
44	0.00
45	0.00
46	0.00
47	0.00
48	0.00
49	0.00
50	0.00
51	0.00
52	0.00
53	0.00
54	0.00
55	0.00
56	0.00
57	0.00
58	0.00
59	0.00
60	0.00
61	0.00
62	0.00
63	0.00
64	0.00
65	0.00
66	0.00
67	0.00
68	0.00
69	0.00
70	0.00
71	0.00
72	0.00
73	0.00
74	0.00
75	0.00
76	0.00
77	0.00
78	0.00
79	0.00
80	0.00
81	0.00
82	0.00
83	0.00
84	0.00
85	0.00
86	0.00
87	0.00
88	0.00
89	0.00
90	0.00
91	0.00
92	0.00
93	0.00
94	0.00
95	0.00
96	0.00
97	0.00
98	0.00
99	0.00

Matches 62; Conserva.

```
Query Match      8.1%; Score 29; DB 3; Length 2158;
Best Local Similarity 53.0%; Pred. No. 1.8;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
```

RESULT 10  
PCT-US95-12724-1/c

NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.

RESULT 11  
US-08-698-

US-08-698-551-11/C  
; Sequence 11, Application US/08698551  
; Patent No. 5712381

```

; GENERAL INFORMATION:
: APPLICANT: Lin. Lib-Ling

```

APPLICANT: Chen, Jennifer H.

APPLICANT: Schievella, Andrea  
APPLICANT: Graham, James

;	TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
:	TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING

; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

Query Match 8.1%; Score 29; DB 6; Length 2158;

Best Local Similarity 53.0%; Pred. No. 1.8;  
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

0v 195 tottaagcaagatcagcacaaagcccaaacggcaagcgtcacattggcctgtttaagccaac 254

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

DD 19ATTTGCTGGCTGTATACACAAAGCCCCCAATGTGCGACATTTCCCAATAAGGCGCCCTACACCTTCT

Qy 255 agctgtggcgctccacgtaccacitctctaaaaaacagggaagcactggagccagcgca 311



```
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1822
; US-08-494-440B-11

Query Match      8.1%; Score 29; DB 3; Length 2735;
Best Local Similarity 53.0%; Pred. No. 2;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 195 tgttaagcaagatgagcagaagcccaaacggaagcgtcacattggctgttaagccaac 254
      || || || || || || || || || || || || || || || || || || || ||
DB 1121 TGATTTCGCTGGCTGTACACAAGCCCAATGTGCGACTTTCCCATAGGCGCCTCACCTTCT 1062
      || || || || || || || || || || || || || || || || || || || ||
QY 255 agctgtggcgtccacgtaccacttctctctctctctctctctctctctctctctctctct 311
      || || || || || || || || || || || || || || || || || || || ||
DB 1061 TCGGATGTCATCTTATTACCTTCATCAGCAGCATGTAGGATGAGGTTGTGCA 1005
      || || || || || || || || || || || || || || || || || || || ||

RESULT 14
US-08-533-901B-11/c
; Sequence 11, Application US/08533901B
; Patent No. 5852173
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,901B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1822
; US-08-533-901B-11
```

```
Query Match      8.1%; Score 29; DB 3; Length 2735;
Best Local Similarity 53.0%; Pred. No. 2;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 195 tgttaagcaagatgagcagaagcccaaacggaagcgtcacattggcgtgttaagccaac 254
      || || || || || || || || || || || || || || || || || || || ||
DB 1121 TGATTTCGCTGGCTGTACACAAGCCCAATGTGCGACTTTCCCATAGGCGCCTCACCTTCT 1062
      || || || || || || || || || || || || || || || || || || || ||
QY 255 agctgtggcgtccacgtaccacttctctctctctctctctctctctctctctctctctct 311
      || || || || || || || || || || || || || || || || || || || ||
DB 1061 TCGGATGTCATCTTATTACCTTCATCAGCAGCATGTAGGATGAGGTTGTGCA 1005
      || || || || || || || || || || || || || || || || || || || ||

RESULT 15
US-08-839-032A-11/c
; Sequence 11, Application US/08839032A
; Patent No. 5891675
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,032A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232DDIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1822
; US-08-839-032A-11
```

```
Query Match      8.1%; Score 29; DB 3; Length 2735;
Best Local Similarity 53.0%; Pred. No. 2;
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 195 tgttaagcaagatgagcagaagcccaaacggaagcgtcacattggcgtgttaagccaac 254
      || || || || || || || || || || || || || || || || || || || ||
DB 1121 TGATTTCGCTGGCTGTACACAAGCCCAATGTGCGACTTTCCCATAGGCGCCTCACCTTCT 1062
      || || || || || || || || || || || || || || || || || || || ||
QY 255 agctgtggcgtccacgtaccacttctctctctctctctctctctctctctctctctctct 311
      || || || || || || || || || || || || || || || || || || || ||
DB 1061 TCGGATGTCATCTTATTACCTTCATCAGCAGCATGTAGGATGAGGTTGTGCA 1005
      || || || || || || || || || || || || || || || || || || || ||
```

Search completed: December 18, 2000, 14:41:09  
Job time: 16836 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 38.84 Seconds  
(without alignments)  
3481.940 Million cell updates/sec

Title: US-09-301-906-4

Perfect score: 360

Sequence: 1 gtcagcggtcagtcagcgc.....aggcgcgttgacgtttggc 360

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_36:\*  
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	360	100.0	360	21	249203
2	360	100.0	6714	21	249208
3	360	100.0	17919	21	249200
C 4	33	9.2	1178	14	Q48025
C 5	32.2	8.9	6714	21	249208
C 6	32.2	8.9	17919	21	249200
C 7	31	8.6	997	14	Q48024
C 8	30.8	8.6	1149	21	253468
C 9	30.4	8.4	567	21	A29550
C 10	30.2	8.4	1105	20	252903
C 11	29.8	8.3	360	21	249203
C 12	29.4	8.2	807	16	Q91362

13	29.4	8.2	807	19	V01408
C 14	29.4	8.2	3385	21	295275
C 15	29.4	8.2	11808	19	V33433
C 16	29.4	8.2	11994	17	T43075
C 17	29.2	8.1	696	21	253906
C 18	29.2	8.1	696	21	253907
C 19	29.2	8.1	696	21	253908
C 20	29.2	8.1	849	21	253909
C 21	29.2	8.1	849	21	253910
C 22	29.2	8.1	849	21	253911
C 23	29.2	8.1	1149	21	253466
C 24	29.2	8.1	1149	21	253467
C 25	29.2	8.1	5059	20	X84332
C 26	29.2	8.1	5338	20	X13242
C 27	29	8.1	1560	16	Q83201
C 28	29	8.1	2158	17	T15228
C 29	29	8.1	2158	18	T94631
C 30	29	8.1	2735	17	T15233
C 31	29	8.1	2735	18	T94636
C 32	29	8.1	6002	18	T94638
C 33	29	8.1	7360	17	T43072
C 34	29	8.1	7360	19	V33410
C 35	28.6	7.9	910715	20	X20248
C 36	28.4	7.9	5360	20	X13155
C 37	28.2	7.8	300	21	A00658
C 38	28.2	7.8	864	7	M60488
C 39	28.2	7.8	864	11	Q03317
C 40	28.2	7.8	1094	19	T98728
C 41	28.2	7.8	6567	11	Q03324
C 42	28	7.8	435	19	V29023
C 43	28	7.8	463	18	X83249
C 44	28	7.8	463	19	V68847
C 45	28	7.8	1689	14	Q52498

#### ALIGNMENTS

RESULT 1

249203

ID 249203 standard; DNA; 360 BP.

XX

AC 249203;

XX

DT 07-MAR-2000 (first entry)

XX

DE GLRAV-3 polyprotein proteinase domain DNA.

XX

KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;

KW viral disease; yield loss; sugar content; inhibition; infection;

KW replication; polyprotein; domain; proteinase; methyltransferase;

KW helicase; RNA-dependent; RNA polymerase; untranslated region;

KW transgenic plant; component; resistant; truncation; deletion; antisense;

XX expression; detection; antibody; ds.

OS Grapevine leafroll-associated virus 3.

XX

FH Key Location/Qualifiers

FT mat\_peptide 1..360

FT /tag- a

FT /product= "GLRAV-3 polyprotein proteinase domain"

XX

XX WO9955880-A1.

PN

XX

PD 04-NOV-1999.

XX

XX 29-APR-1999; 99WO-US09307.

PF

XX

PR 29-APR-1998; 98US-0083404.

XX

PA (CORR ) CORNELL RES FOUND INC.

XX

PI Gonsalves D, Ling K;

```

XX DR WPI: 2000-062035/05.
XX P-PSDB: Y58143.
XX PT Newly isolated grapevine leafroll virus protein or polypeptide useful
XX PT for producing transgenic plants conferring viral disease resistance -
XX PS Claim 12: Fig 5; 84pp; English.
XX
XX This sequence represents DNA encoding the proteinase domain of the
XX polyprotein (Y58148) from the grapevine leafroll-associated virus 3
XX (GLRAV-3). Leafroll is a serious viral disease, occurring wherever
XX grapes are grown. Although not lethal, it causes yield losses and
XX reduction in sugar content. The virus encodes several proteins,
XX which may serve as targets for the inhibition of viral infection or
XX replication. These proteins include the 242-248 kD polyprotein (Y58148),
XX encoded by open reading frame (ORF) 1a, and which comprises a proteinase
XX domain (Y58143), a methyltransferase domain (Y58144) and a helicase
XX domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
XX 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
XX Nucleotides encoding these proteins, or fragments thereof, and the 5',
XX - and 3' untranslated regions (UTRs) of the genome (249201-249202) are
XX useful for the generation of transgenic plants and plant components.
XX Such transgenic plants may be resistant to viral disease, for example,
XX this property being conferred on the plants via the use of nucleotides
XX encoding truncated or internally deleted proteins, or via the use of
XX antisense nucleotides to inhibit viral gene expression. The nucleotides
XX may additionally be used in the detection of viral nucleic acids in a
XX tissue sample. The proteins can be used to generate antibodies which
XX may be used to detect GLRAV-3 in plant samples. The isolation of
XX GLRAV-3 facilitates the production of agents that reduce the risk of
XX infection or damage by the virus in vineyards.
XX
XX Sequence 360 BP; 90 A; 97 C; 100 G; 73 T; 0 other;

Query Match      100.0%; Score: 360; DB 21; Length 360;
Best Local Similarity 100.0%; Pred. NO. 1.6e-112;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcagcggtcagtcagcggtgaggggagtgatggtgaagaggtcttgatgagggcaagg 60
Db 1 gtcagcggtcagtcagcggtgaggggagtgatggtgaagaggtcttgatgagggcaagg 60
QY 61 acctcaacttcgcaacttcgacgtgctgatttcgacgtcgtattcgaagctgtttct 120
Db 61 acctcaacttcgcaacttcgacgtgctgatttcgacgtcgtattcgaagctgtttct 120
QY 121 aatgcattacttgcgtacactacacgcgggtagtcgcgtatgcccgctcaagcgcgag 180
Db 121 aatgcattacttgcgtacactacacgcgggtagtcgcgtatgcccgctcaagcgcgag 180
QY 181 cagcctaaccggtgcttaagcaagatgagcagaagcccaacgcaagcgtcacattgg 240
Db 181 cagcctaaccggtgcttaagcaagatgagcagaagcccaacgcaagcgtcacattgg 240
QY 241 gctgttaagcaacagcgtgttgctccacgtaccacttcctataaaacagaagcactg 300
Db 241 gctgttaagcaacagcgtgttgctccacgtaccacttcctataaaacagaagcactg 300
QY 301 gagccagcgaatcagtcaccacacagtcgcttgaggagagagcgccgttgacgtttggc 360
Db 301 gagccagcgaatcagtcaccacacagtcgcttgaggagagagcgccgttgacgtttggc 360

RESULT 2
ID 249208
XX 249208 standard; DNA; 6714 BP.
XX AC 249208;
XX DT 07-MAR-2000 (first entry)
XX

```

```

DE XX GLRAV-3 ORF 1a, encoding polyprotein.
KW XX Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRAV-3;
KW XX viral disease; yield loss; sugar content; inhibition; infection;
KW XX replication; polyprotein; domain; proteinase; methyltransferase;
KW XX helicase; RNA-dependent; RNA polymerase; untranslated region;
KW XX transgenic plant; component; resistant; truncation; deletion; antisense;
KW XX expression; detection; antibody; ds.
XX
XX Grapevine leafroll-associated virus 3.
XX
XX Key Location/Qualifiers
XX CDS 1..6714
XX /note= "ORF 1a"
XX /product= "GLRAV-3 polyprotein"
XX /tag= a
XX mat_peptide 253..612
XX /product= "GLRAV-3 polyprotein proteinase domain
XX (Y58143)"
XX /tag= b
XX mat_peptide 1378..2193
XX /product= "GLRAV-3 polyprotein methyltransferase domain
XX (Y58144)"
XX /tag= c
XX mat_peptide 5764..6636
XX /product= "GLRAV-3 polyprotein helicase domain (Y58145)"
XX /tag= d
XX W09955880-A1.
XX
XX 04-NOV-1999.
XX
XX 29-APR-1999; 99WO-US09307.
XX
XX 29-APR-1998; 98US-0083404.
XX
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Gonsalves D, Ling K;
XX
XX WPI: 2000-062035/05.
XX P-PSDB: Y58148.
XX
XX Newly isolated grapevine leafroll virus protein or polypeptide useful
XX PT for producing transgenic plants conferring viral disease resistance -
XX PS Claim 7; Fig 4; 84pp; English.
XX
XX This sequence represents open reading frame (ORF) 1a which encodes
XX the polyprotein of the grapevine leafroll-associated virus 3
XX (GLRAV-3). Leafroll is a serious viral disease, occurring wherever
XX grapes are grown. Although not lethal, it causes yield losses and
XX reduction in sugar content. The virus encodes several proteins,
XX which may serve as targets for the inhibition of viral infection or
XX replication. These proteins include the 242-248 kD polyprotein (Y58148),
XX encoded by open reading frame (ORF) 1a, and which comprises a proteinase
XX domain (Y58143), a methyltransferase domain (Y58144) and a helicase
XX domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
XX 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
XX Nucleotides encoding these proteins, or fragments thereof, and the 5',
XX - and 3' untranslated regions (UTRs) of the genome (249201-249202) are
XX useful for the generation of transgenic plants and plant components.
XX Such transgenic plants may be resistant to viral disease, for example,
XX this property being conferred on the plants via the use of nucleotides
XX encoding truncated or internally deleted proteins, or via the use of
XX antisense nucleotides to inhibit viral gene expression. The nucleotides
XX may additionally be used in the detection of viral nucleic acids in a
XX tissue sample. The proteins can be used to generate antibodies which
XX may be used to detect GLRAV-3 in plant samples. The isolation of
XX GLRAV-3 facilitates the production of agents that reduce the risk of
XX infection or damage by the virus in vineyards.
XX
XX Sequence 6714 BP; 1675 A; 1429 C; 1877 G; 1733 T; 0 other;

```

Query Match		100.0%;	Score 360;	DB 21;	Length 6714;
Best Local Similarity		100.0%;	Pred. No. 6.1e-112;		
Matches 360;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	gtcagcgctcagtcagcgcgtgagaggggtagtgaagaggtcttgatggagggcaag	60		
Db	253	gtcagcgctcagtcagcgcgtgagaggggtagtgaagaggtcttgatggagggcaag	312		
QY	61	acctcaacttcgcaacttcgacgtgctgtgatttcgacgtcgatttcgaagctgtttct	120		
Db	313	acctcaacttcgcaacttcgacgtgctgtgatttcgacgtcgatttcgaagctgtttct	372		
QY	121	aatgcattacttgcgtacactaccacgggtagtcgctatgccccgcgtcaagcgcgag	180		
Db	373	aatgcattacttgcgtacactaccacgggtagtcgctatgccccgcgtcaagcgcgag	432		
QY	181	cagcctaaacggcgtgttaagcaagatgagcagaagcccaaacggcgaagcgtcacattg	240		
Db	433	cagcctaaacggcgtgttaagcaagatgagcagaagcccaaacggcgaagcgtcacattg	492		
QY	241	gctgttaagccaacagctgttgccgtccacgtaccacttccctaaataaacagggaagcactg	300		
Db	493	gctgttaagccaacagctgttgccgtccacgtaccacttccctaaataaacagggaagcactg	552		
QY	301	gagccagcgaatcagtcgcccaacacagctgttgaggagagaagccgccttgacgtttggc	360		
Db	553	gagccagcgaatcagtcgcccaacacagctgttgaggagagaagccgccttgacgtttggc	612		
RESULT 3					
Z49200					
ID	Z49200 standard; DNA; 17919 BP.				
XX					
AC	Z49200;				
XX					
DT	07-MAR-2000 (first entry)				
DE	Grapevine leafroll-associated virus 3 (GLRaV-3) genome.				
XX					
KW	Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;				
KW	viral disease; yield loss; sugar content; inhibition; infection;				
KW	replication; polyprotein; domain; proteinase; methyltransferase;				
KW	helicase; RNA-dependent; RNA polymerase; untranslated region;				
KW	transgenic plant; component; resistant; truncation; deletion; antisense;				
KW	expression; detection; antibody; ds.				
XX					
OS	Grapevine leafroll-associated virus 3.				
XX					
Key	Location/Qualifiers				
5'UTR	1..158				
FT	/*tag= a				
FT	159..6872				
FT	/*tag= b				
FT	/*note= "ORF 1a"				
FT	/*product= "GLRaV-3 polyprotein (Y58148)"				
FT	411..770				
FT	/*tag= c				
FT	/*product= "GLRaV-3 polyprotein proteinase domain (Y58143)"				
FT	1536..2351				
FT	/*tag= d				
FT	/*product= "GLRaV-3 polyprotein methyltransferase domain (Y58144)"				
FT	5922..6794				
FT	/*tag= e				
FT	/*product= "GLRaV-3 polyprotein helicase domain (Y58145)"				
FT	6877..8478				
FT	/*tag= f				
FT	/*note= "ORF 1b"				
FT	/*product= "GLRaV-3 RNA-dependent RNA polymerase (Y58146)"				
FT	8708..8863				
FT	CDS				

FT	/*tag= g				
FT	/*note= "ORF 2"				
FT	9930..10067				
FT	/*tag= h				
FT	/*note= "ORF 3"				
FT	10086..11735				
FT	/*tag= i				
FT	/*note= "ORF 4"				
FT	11728..13179				
FT	/*tag= j				
FT	/*note= "ORF 5"				
FT	13269..14210				
FT	/*tag= k				
FT	/*note= "ORF 6"				
FT	14273..15706				
FT	/*tag= l				
FT	/*note= "ORF 7"				
FT	15717..16274				
FT	/*tag= m				
FT	/*note= "ORF 8"				
FT	16271..16804				
FT	/*tag= n				
FT	/*note= "ORF 9"				
FT	16811..17350				
FT	/*tag= o				
FT	/*note= "ORF 10"				
FT	17353..17463				
FT	/*tag= p				
FT	/*note= "ORF 11"				
FT	/*product= "GLRaV-3 ORF 11-encoded protein (Y58147)"				
FT	17460..17642				
FT	/*tag= q				
FT	/*note= "ORF 12"				
FT	17643..17919				
FT	/*tag= r				
XX	W09955880-A1.				
XX	PD	04-NOV-1999.			
XX	PF	29-APR-1999; 99WO-US09307.			
XX	PR	29-APR-1998; 98US-0083404.			
XX	PA	(CORR ) CORNELL RES FOUND INC.			
XX	PI	Gonsalves D, Ling K;			
XX	DR	WPI; 2000-062035/05.			
XX	DR	P-PSDB; Y58143, Y58144, Y58145, Y58146, Y58147, Y58148.			
XX	PT	Newly isolated grapevine leafroll virus protein or polypeptide useful for producing transgenic plants conferring viral disease resistance -			
XX	PS	Example 1; Fig 2; 84pp; English.			
XX	CC	This sequence represents the genome of grapevine leafroll-associated virus 3 (GLRaV-3). Leafroll is a serious viral disease, occurring wherever grapes are grown. Although not lethal, it causes yield losses and reduction in sugar content. The virus encodes several proteins, which may serve as targets for the inhibition of viral infection or replication. These proteins include a 242-248 kD polyprotein (Y58148), encoded by open reading frame (ORF) 1a, and which comprises a proteinase domain (Y58143), a methyltransferase domain (Y58144) and a helicase domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF 1b; and a protein of unspecified function (Y58147), encoded by ORF 11. Nucleotides encoding these proteins, or fragments thereof, and the 5' and 3' untranslated regions (UTRs) of the genome (24201-24920) are useful for the generation of transgenic plants and plant components. Such transgenic plants may be resistant to viral disease, for example, this property being conferred on the plants via the use of nucleotides encoding truncated or internally deleted proteins, or via the use of antisense nucleotides to inhibit viral gene expression. The nucleotides			





```

XX PA (CORR ) CORNELL RES FOUND INC.
XX PI Gonsalves D, Ling K;
XX PF WPI: 2000-062035/05
XX DR P-PSDB: Y58143, Y58144, Y58145, Y58146, Y58147, Y58148.
XX DR Newly isolated grapevine leafroll virus protein or polypeptide useful
XX PT for producing transgenic plants conferring viral disease resistance -
XX PS Example 1; Fig 2; 84pp; English.
XX
CC This sequence represents the genome of grapevine leafroll-associated
CC virus 3 (GLRAV-3). Leafroll is a serious viral disease, occurring
CC wherever grapes are grown. Although not lethal, it causes yield losses
CC and reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include a 242-248 kD polyprotein (Y58148),
CC domain (Y58143), a methyltransferase domain (Y58144), and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5',
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRAV-3 in plant samples. The isolation of
CC GLRAV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX SQ Sequence 17919 BP; 4729 A; 3527 C; 4740 G; 4923 T; 0 other;

Query Match      8.9%; Score 32.2; DB 21; Length 17919;
Best Local Similarity 51.8%; Pred. No. 1.4;
Matches 73; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 220 aaacgcgaagcgtcacattggcgtttaagccacagctgttgccacgtaccactt 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 774 AAGGCCCAACAGCTCAAGCGCGCCTTCCTCCACGACTGTGTGGGACTGATTCGGTG 715

QY 280 cctaaaaaacaggacactggagccagcgcaatcagtcctccacacagctgttgaggag 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 GCTCCAGTGTCTCTCTGTTTATTAGGAAGTGTGACGTGGACGCCAACACGCTGTGGCTTAA 655

QY 340 aagggccgcttgacgtttggc 360
   || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 CAGCCCAATGTGACGCTGCC 634

RESULT 7
Q48024/c
ID Q48024 standard; DNA; 997 BP.
XX AC Q48024;
XX DT 21-APR-1994 (first entry)
XX DE AIGF gene #1.
XX DE Cellular growth factor; androgen-induced growth factor; AIGF;
KW cell growth; reagent; detection; anticancer; ss.
XX OS Homo sapiens.
XX FH Key
XX CDS 174..821
Location/Qualifiers
FT

```

```

FT
FT
XX
XX
PN JP05222096-A.
XX
XX 31-AUG-1993.
XX
XX 28-JAN-1992; 92JP-0038717.
XX
XX 28-JAN-1992; 92JP-0038717.
XX
XX (SHIO ) SHIONOGI & CO LTD.
XX
XX WPI: 1993-309194/39.
XX DR P-PSDB; R41690.
XX
XX Cell growth factor useful for detection of anticancer substance -
XX PT has specified amino acid sequence
XX PS Claim 1; Page 6-7 ; 10pp; Japanese.
XX
XX The sequences given in 048024-25 encode cellular growth factors, esp.
XX androgen-induced growth factor (AIGF). The pref. protein of the
XX invention comprises amino acids His35-Arg215 of sequence #1 or Gln23-
XX Arg268 of sequence #2. The AIGFs encoded by these sequences are
XX useful as cell growth reagents and for the detection of anticancer
XX substances.
XX SQ Sequence 997 BP; 192 A; 330 C; 293 G; 182 T; 0 other;

Query Match      8.6%; Score 31; DB 14; Length 997;
Best Local Similarity 52.8%; Pred. No. 0.92;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 220 aaacgcgaagcgtcacattggcgtttaagccacagctgttgccacgtaccactt 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 AACAGCAGAGTGCACAGCAGCAGCAGCTCAGCGGGAGCGGGGCTGCCCATGGCGCGCGG 165

QY 280 cctaaaaaacaggacactggagccagcgcaatcagtcctccacacagctgttgaggag 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 CCCCAGGAGCAGTGTAGAGCCCGAGCGGGGTACGCGGTCCCACTGGAGGCCGCGGGGGA 105

QY 340 aagggcg 346
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 CAAGCCG 98

RESULT 8
Z53468/c
ID Z53468 standard; DNA; 1149 BP.
XX AC Z53468;
XX DT 21-MAR-2000 (first entry)
XX DE Neisseria meningitidis ORF 237 partial DNA sequence SEQ ID NO:885.
XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy; ds.
XX OS Neisseria meningitidis.
XX PN WO9957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.

```

PR	02-SEP-1998;	98US-009906Z.	
PR	09-OCT-1998;	98US-0103749.	
PR	09-OCT-1998;	98US-0103794.	
PR	09-OCT-1998;	98US-0103796.	
PR	25-FEB-1999;	99US-0121528.	
XX			
PA	(CHIR ) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
XX			
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;		
PI	Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;		
PI	Tettelin H, Venter JC;		
XX			
DR	WPI: 2000-062150/05.		
DR	P-PSDB; Y74706.		
XX			
PT	Novel Neisserial polypeptides predicted to be useful antigens for		
PT	vaccines and diagnostics		
XX			
PS	Claim 7; Page 542; 1453pp; English.		
XX			
CC	253015 to 254536, 254577 to 254615, and Y74253 to Y75941 represent		
CC	novel <i>Neisseria meningitis</i> and <i>N. gonorrhoeae</i> polynucleotides and		
CC	polypeptides. 254537 to 254576 and 254616 to 255473 represent PCR		
CC	primers used in the exemplification of the present invention. The		
CC	polypeptides, the polynucleotides, antibodies and compositions of		
CC	the invention can be used as vaccines, as diagnostic reagents, and as		
CC	immunogenic compositions. The polypeptides can be used in the		
CC	manufacture of medicaments for treating or preventing infection due to		
CC	<i>Neisserial bacteria</i> (e.g. <i>meningitis</i> and <i>septicaemia</i> ), to detect the		
CC	presence of <i>Neisseria bacteria</i> , or to raise antibodies. They may also		
CC	be used to screen for agonists or antagonists, which may themselves		
CC	have use as antibacterial agents. The polynucleotides of the invention		
CC	may also be used in gene therapy protocols.		
XX			
SQ	Sequence 1149 BP; 288 A; 297 C; 286 G; 278 T; 0 other;		

```

Query Match      8.6%; Score 30.8; DB 21; Length 1149;
Best Local Similarity 70.7%; Pred. No. 1.1;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY . _282 taaaaacaggaagcactggagccagcgcaatcagtcaccacaacagtcgttgaggagag 339
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   802 TGA AAAACAGCTGGTTTGGAGCAAGCGCAATGTTTTTGCCAAACAGTCTTTGCAGGCG 745

RESULT          9
A29550
ID A29550 standard; DNA; 567 BP.
XX
AC A29550;
XX
XX
DT 09-AUG-2000 (first entry)
XX
DE HIV codon altered env nucleotide sequence #2.
XX
KW Erythropoietin; EPO; G-CSF; granulocyte colony stimulating factor;
wobble; codon altered gene; shuffling; modification; vaccine;
KW Insulin; peptide hormone; growth factor; cytokine; interferon;
KW Interleukin; leukemia inhibitory factor; oncostatin M;
KW transcription activator; expression activator; infectious organism; ds.
XX
OS Human immunodeficiency virus type 1.
Synthetic.
OS
XX
PN WO200018906-A2.
XX
PD 06-APR-2000.
XX
XX
PF 28-SEP-1999; 99WO-US22588.
XX
PR 29-SEP-1998; 98US-0102362.

```

```

PR 29-JAN-1999; 99US-0117729.
PR 05-FEB-1999; 99US-0118813.
PR 24-JUN-1999; 99US-0141049.
XX (MAXY-) MAXYGEN INC.
XX
XX Patten PA, Liu L, Stemmer WPC;
XX
XX WPI; 2000-303449/26.
XX
XX Novel methods for recombining codon-altered libraries of nucleic acids
PT used to produce new proteins and new vectors with reduced rates of
PT reversion to wild type
XX
XX Example; Fig 18A; 92pp; English.
XX
XX A method has been developed of making codon altered nucleic acids (NAS),
CC comprising providing a NA sequence (NA) which encodes a polypeptide
CC (PI), providing codon altered NA sequences, each encoding PI or a.
CC modified form of it, and recombining the codon altered NA sequences to
CC produce a target codon altered NA which encodes a second protein.
CC The method of the invention can be used for recombining codon-altered
CC libraries of nucleic acids to produce new proteins, which have
CC improvements in a desirable characteristic. Target nucleic acids
CC include those coding for therapeutic proteins such as erythropoietin
CC (EPO), insulin, peptide hormones, growth factors, cytokines, interferons,
CC interleukins, leukaemia inhibitory factor, and oncostatin M, as well as
CC transcription and expression activators and proteins from infectious
CC organisms for use as vaccines. The method can also be used to produce
CC attenuated viruses which have reduced rates of reversion to wild type.
CC The present sequence represents an HIV codon altered env nucleotide
CC sequence, which is used in an example from the present invention.
XX
XX Sequence 567 BP; 32 A; 129 C; 222 G; 184 T; 0 Other;
XX
XX
XX Query Match 8.4%; Score 30.4; DB 21; Length 567;
XX Best Local Similarity 49.4%; Pred. No. 1.1;
XX Matches 79; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
XX
XX Qy 68 cttccgaacttcgagcgtgtctgatttcgacgtcgatttcgaagctgtttctaatgcat 127
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 376 ctgctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcg 435
XX
XX Qy 128 tactctgcgtacataccaccggtagtcgcgtatgccccgtcaagcgcgagcagccta 187
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 436 tcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 495
XX
XX Qy 188 aaccggctgttaagcaagatgagcagaagcccaacaggca 227
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX Db 496 agcagcagcagcagcagcagcagcagcagcagcagcagcagca 535
XX
XX RESULT 10
XX Z52903/C
XX ID Z52903 standard; cDNA; 1105 BP.
XX
XX AC Z52903;
XX
XX 14-MAR-2000 (first entry)
XX
XX Human prostate tumor cDNA library derived EST fragment #46.
XX
XX Pancreas; tumor; ESP; expressed sequence tag; human; cytostatic;
KW treatment; ds.
XX
XX Homo sapiens.
XX
XX DE19820190-A1.
XX
XX 04-NOV-1999.
XX
XX 28-APR-1998; 98DE-1020190.
XX

```

```
XX 28-APR-1998; 98DE-1020190.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI; 1999-621386/54.
XX P-PSDB; Y73950, Y73951, Y73952.
XX
XX New human nucleic acid sequences from pancreatic tumors, and related
XX proteins -
XX
XX Claim 2; Page 219-220; 502pp; German.
XX
XX This invention describes novel polypeptides and their encoding nucleic
XX acids derived from human pancreatic tumor tissue which have cytostatic
XX activity. The sequences are also useful in producing pharmaceutical
XX compositions for treatment of pancreatic tumors. 252858-253014 represent
XX expressed sequence tag (EST) fragments derived from a human pancreatic
XX tumor cDNA library and which encode the proteins represented in
XX - Y73814-Y74252.
XX
XX Sequence 1105 BP; 275 A; 299 C; 328 G; 203 T; 0 other;

Query Match      8.4%; Score 30.2; DB 20; Length 1105;
Best Local Similarity 53.9%; Pred. No. 1.8;
Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 106 ttcgagctgtttcgaattcattctgtcgtacactaccacgggtagtgcgtatgcc 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1100 TTCAACAGGTAACCTAATAAGCTGTGCGCCCTGCGGCCCCACCGTGAATACGCC 1041

QY 166 ccgcgcagcgcgagcagcctaaacccgctgttaagcaagatgacgagccca 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1040 CCAGTGGGTGAACACAGGCCCACTGTGGGTATTGATGAACATGGGCACCAAGCCCA 986

RESULT 11
ID 249203 standard; DNA; 360 BP.
XX
AC 249203;
XX
DT 07-MAR-2000 (first entry)
XX
DE GLRaV-3 polyprotein proteinase domain DNA.
XX
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody; ds.
XX
OS Grapevine leafroll-associated virus 3.
XX
FH Key Location/Qualifiers
FT mat_peptide 1..360
FT /tag= a
FT /product= "GLRaV-3 polyprotein proteinase domain"
XX
PN W09955880-A1.
XX
XX 04-NOV-1999.
XX
XX 29-APR-1999; 99WO-US09307.
XX
XX 29-APR-1998; 98US-0083404.
XX (CORR ) CORNELL RES FOUND INC.
XX
```

```
PI Gonsalves D, Ling K;
XX
XX WPI; 2000-062035/05.
XX P-PSDB; Y58143.
XX
XX Newly isolated grapevine leafroll virus protein or polypeptide useful
XX for producing transgenic plants conferring viral disease resistance -
XX
XX Claim 12; Fig 5; 84pp; English.
XX
XX This sequence represents DNA encoding the proteinase domain of the
XX polyprotein (Y58148) from the grapevine leafroll-associated virus 3
XX (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
XX grapes are grown. Although not lethal, it causes yield losses and
XX reduction in sugar content. The virus encodes several proteins,
XX which may serve as targets for the inhibition of viral infection or
XX replication. These proteins include the 242-248 kD polyprotein (Y58148),
XX encoded by open reading frame (ORF) 1a, and which comprises a proteinase
XX domain (Y58143), a methyltransferase domain (Y58144) and a helicase
XX domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
XX 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
XX Nucleotides encoding these proteins, or fragments thereof, and the 5',
XX and 3' untranslated regions (UTRs) of the genome (249201-249202) are
XX useful for the generation of transgenic plants and plant components.
XX Such transgenic plants may be resistant to viral disease, for example,
XX this property being conferred on the plants via the use of nucleotides
XX encoding truncated or internally deleted proteins, or via the use of
XX antisense nucleotides to inhibit viral gene expression. The nucleotides
XX may additionally be used in the detection of viral nucleic acids in a
XX tissue sample. The proteins can be used to generate antibodies which
XX may be used to detect GLRaV-3 in plant samples. The isolation of
XX GLRaV-3 facilitates the production of agents that reduce the risk of
XX infection or damage by the virus in vineyards.
XX
XX Sequence 360 BP; 90 A; 97 C; 100 G; 73 T; 0 other;

Query Match      8.3%; Score 29.8; DB 21; Length 360;
Best Local Similarity 51.1%; Pred. No. 1.5;
Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 224 gccaagcgtcacattgggtgttaagccaacagctgtgtgctccacgtaccacttcta 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 360 GCCAACAGCTCAGGCGCGCTTCTCTCCACACGACTGTGTGGGACTGATTGCGGTGGCTC 301

QY 284 aaaaacaggagcactggagccagcgcaatcagtcacacacagctgttgagagagaagg 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 300 CAGTGTCTTCTGTGTTTATTAGGAAGTGTGACGTGGACGCCAACAGCTGTGGCTTAACAGC 241

QY 344 ccgccttgacgtttggc 360
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 240 CCAATGTGACGCTTGCC 224

RESULT 12
QY1362
ID QY1362 standard; cDNA; 807 BP.
XX
AC QY1362;
XX
DT 24-NOV-1995 (first entry)
XX
DE T-cell receptor alpha chain (TCR alpha) from hybridoma 3B3.
XX
KW T-cell receptor alpha chain; TCR alpha; hyperimmune;
KW immunodeficiency; Cd4+; helper T cell; hybridoma 3B3; ss.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..804
FT /tag= a
XX
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 18, 2000, 10:00:33 ; Search time 856.91 Seconds  
(without alignments)  
2597.484 Million cell updates/sec

Title: US-09-301-906-4  
Perfect score: 360  
Sequence: 1 gtcagcggtctcagtcagcgc.....aggcgcgcttgacgtttggc 360

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*  
12: gb\_est12.\*  
13: gb\_est13.\*  
14: gb\_est14.\*  
15: gb\_est15.\*  
16: gb\_est16.\*  
17: gb\_est17.\*  
18: gb\_est18.\*  
19: gb\_est19.\*  
20: gb\_est20.\*  
21: gb\_est21.\*  
22: gb\_est22.\*  
23: gb\_est23.\*  
24: gb\_est24.\*  
25: gb\_est25.\*  
26: gb\_est26.\*  
27: gb\_est27.\*  
28: gb\_est28.\*  
29: gb\_est29.\*  
30: gb\_est30.\*  
31: gb\_est31.\*  
32: gb\_est32.\*  
33: gb\_est33.\*  
34: gb\_est34.\*  
35: gb\_est35.\*  
36: gb\_est36.\*  
37: gb\_est37.\*  
38: gb\_est38.\*  
39: gb\_est39.\*  
40: gb\_est40.\*  
41: em\_estba.\*  
42: em\_estfun.\*  
43: em\_esthum1.\*

44: em\_esthum2.\*  
45: em\_esthum3.\*  
46: em\_esthum4.\*  
47: em\_esthum5.\*  
48: em\_esthum6.\*  
49: em\_esthum7.\*  
50: em\_esthum8.\*  
51: em\_esthum9.\*  
52: em\_esthum10.\*  
53: em\_esthum11.\*  
54: em\_esthum12.\*  
55: em\_esthum13.\*  
56: em\_esthum14.\*  
57: em\_esthum15.\*  
58: em\_esthum16.\*  
59: em\_esthum17.\*  
60: em\_esthum18.\*  
61: em\_esthum19.\*  
62: em\_esthum20.\*  
63: em\_estin1.\*  
64: em\_estin2.\*  
65: em\_estin3.\*  
66: em\_estin4.\*  
67: em\_estom.\*  
68: em\_estov1.\*  
69: em\_estov2.\*  
70: em\_estpl1.\*  
71: em\_estpl2.\*  
72: em\_estpl3.\*  
73: em\_estpl4.\*  
74: em\_estpl5.\*  
75: em\_estrol.\*  
76: em\_estro2.\*  
77: em\_estro3.\*  
78: em\_estro4.\*  
79: em\_estro5.\*  
80: em\_estro6.\*  
81: em\_estro7.\*  
82: em\_estro8.\*  
83: em\_estro9.\*  
84: em\_estrol1.\*  
85: em\_estrol2.\*  
86: em\_estrol3.\*  
87: gb\_gss1.\*  
88: gb\_gss2.\*  
89: gb\_gss3.\*  
90: gb\_gss4.\*  
91: gb\_gss5.\*  
92: gb\_gss6.\*  
93: gb\_gss7.\*  
94: gb\_gss8.\*  
95: gb\_gss9.\*  
96: gb\_gss10.\*  
97: gb\_gss11.\*  
98: gb\_gss12.\*  
99: gb\_gss13.\*  
100: gb\_gss14.\*  
101: gb\_gss15.\*  
102: gb\_gss16.\*  
103: gb\_gss17.\*  
104: gb\_gss18.\*  
105: gb\_gss19.\*  
106: gb\_gss20.\*  
107: gb\_gss21.\*  
108: gb\_gss22.\*  
109: gb\_gss23.\*  
110: gb\_gss24.\*  
111: gb\_gss25.\*  
112: gb\_gss26.\*  
113: gb\_gss27.\*  
114: gb\_gss28.\*  
115: gb\_gss29.\*  
116: gb\_gss30.\*

117: gb\_gss18:\*  
 118: gb\_gss19:\*  
 119: em\_gss13:\*  
 120: gb\_gss20:\*  
 121: gb\_gss21:\*  
 122: gb\_gss22:\*  
 123: gb\_gss23:\*  
 124: gb\_gss24:\*  
 125: em\_gss14:\*  
 126: em\_gss15:\*  
 127: em\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	36.8	10.2	276	27	BB155217	BB155217
C 2	36.4	10.1	370	40	W99276	W99276 T3079 WVA74
C 3	36.4	10.1	409	40	W69025	W69025 T2568 WVA74
C 4	36.4	10.1	661	100	AQ651482	AQ651482 Sheared D
C 5	35.6	9.9	597	100	AQ652626	AQ652626 Sheared D
C 6	35.2	9.8	511	23	AW659211	AW659211 96177 MAR
C 7	33.8	9.4	609	24	AW765044	AW765044 da87601.Y
C 8	33.4	9.3	341	5	AA696560	AA696560 GM07881.5
C 9	33.4	9.3	634	112	AQ950287	AQ950287 Sheared D
C 10	33.4	9.3	719	35	BE430706	BE430706 SUN007.CO
C 11	33	9.2	331	3	AA334604	AA334604 EST38844
C 12	33	9.2	489	6	AA859269	AA859269 UI-R-E0-C
C 13	33	9.2	713	25	AW955769	AW955769 EST367839
C 14	33	9.2	1101	121	CNS00KK2	AL077673 Drosophila
C 15	32.6	9.1	480	121	CNS00TNV	AL089929 Arabidops
C 16	32.6	9.1	741	36	BE534866	BE534866 601231410
C 17	32.2	8.9	426	36	C95499	C95499 C95499 Citr
C 18	32	8.9	380	23	AW599704	AW599704 ga91908.Y
C 19	32	8.9	400	14	AL361014	AL361014 AL361014
C 20	32	8.9	716	25	AW940039	AW940039 GH01388.3
C 21	32	8.9	1101	121	CNS00BXY	AL057978 Drosophila
C 22	31.8	8.8	421	21	AW248410	AW248410 2820429.5
C 23	31.6	8.8	440	3	AA298086	AA298086 EST113719
C 24	31.6	8.8	457	8	AL109978	AL109978 GH09394.5
C 25	31.6	8.8	625	6	AA736166	AA736166 HL08104.5
C 26	31.4	8.7	424	20	AW233549	AW233549 f37908.x
C 27	31.4	8.7	444	32	BE017212	BE017212 f477h12.Y
C 28	31.4	8.7	551	34	BE201942	BE201942 f103009.x
C 29	31.4	8.7	606	34	BE201791	BE201791 f49807.x
C 30	31.4	8.7	623	36	BE543934	BE543934 601070441
C 31	31.2	8.7	484	19	AV588611	AV588611 AV588611
C 32	31.2	8.7	509	36	BE497852	BE497852 WHE0957_E
C 33	31.2	8.7	555	113	AQ951351	AQ951351 Sheared D
C 34	31.2	8.7	730	121	CNS012WS	AL102118 Drosophila
C 35	31.2	8.7	1049	123	CNS04SBU	AL104995 Tetraodon
C 36	31.2	8.7	1052	122	CNS03BPK	AL236801 Tetraodon
C 37	31	8.6	538	15	AU066586	AU066586 AU066586
C 38	31	8.6	753	40	Z78358	Z78358 HS278358 Hu
C 39	31	8.6	960	34	BE283129	BE283129 601103594
C 40	31	8.6	1006	121	CNS005YL	AL066088 Drosophila
C 41	30.8	8.6	359	35	BE400123	BE400123 AWB011.D0
C 42	30.8	8.6	421	5	AA625956	AA625956 zu92e03.s
C 43	30.8	8.6	457	21	AW342909	AW342909 f480f02.Y
C 44	30.8	8.6	486	36	BE499631	BE499631 WHE0962_C
C 45	30.8	8.6	524	7	AA898051	AA898051 NCC2A1177

## ALIGNMENTS

RESULT 1  
 BB155217/c

# LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

BB155217 276 bp mRNA EST 29-JUN-2000  
 BB155217 RIKEN full-length enriched, 16 days neonate thymus Mus musculus cDNA clone A130024C22 3', mRNA sequence.  
 BB155217  
 BB155217.1 GI:8811147  
 EST.  
 house musculus  
 Mus musculus  
 house musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 276)  
 Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Iizawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toyota,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN MOUSE ESTS (Konno,H., et al.)  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki  
 Genome Exploration Research Group, Life Science Tsukuba Center,  
 Genome Science Laboratory  
 The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: +81-298-36-9013  
 Fax: +81-298-36-9098  
 Email: genome.res@rtc.riken.go.jp,  
 URL:http://genome.rtc.riken.go.jp/  
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

## FEATURES

## source

Location/Qualifiers  
 1..276  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A130024C22"  
 /clone.lib="RIKEN full-length enriched, 16 days neonate thymus"  
 /tissue\_type="thymus"  
 /dev\_stage="16 days neonate"  
 /lab\_host="DH10B"  
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGCAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10, 0 and subtraction to Rot = 185, 0. Second strand cDNA was prepared with the primer adapter of

Db 326 CATACGACCACCCCGGCACATGTTACCGAATGCACCAGCACCTGCTGTACCACC 267





```

/sex="female"
/dev_stage="newly enclosed females: germarium-stage 6"
/lab_host="SOLR"
/notes="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT      73 a   82 c   97 g   89 t
ORIGIN

```

```

Query Match      9.3%; Score 33.4; DB 5; Length 341;
Best Local Similarity 55.7%; Pred. No. 6.2;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 134 tcgtacactaccacgggtagtcggtatgccccctcaagcgcgagcagcctaaacgg 193
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 240 TGGCGGCTACCTCGGTCGGAACGACATCTCGCATAGCGCGCATGGCCAACTGG 181
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 194 ctgttaagcaagatgacagcagcgaacgaacgcgcacattggctgttaa 248
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 180 CTGCTATGGAGGAGGAGCAGCAGGAACCTGGCAACGGAACATTGTGAATTAAA 126
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 9
LOCUS      AQ950287      634 bp      DNA      GSS      27-JAN-2000
DEFINITION Sheared DNA-37L6.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION  AQ950287
VERSION     AQ950287.1 GI:6773552
KEYWORDS   GSS.
SOURCE     Trypanosoma brucei.
ORGANISM   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE  1 (bases 1 to 634)
AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leach,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
            Fraser,C. and Adams,M.
TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
JOURNAL   Unpublished (1999)
COMMENT   Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tldb/mdb/tbdb/.
            Seq primer: M13-Reverse
            Class: Shotgun.

```

```

FEATURES             Location/Qualifiers
     source           1..634
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-37L6"
                     /clone_lib="Sheared DNA"
     notes="Vector: pUC18; Site_1: SmaI; Constructed at the
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
            sheared to give a tight size distribution (approx 2 kb).
            The v + i method used for the library construction is
            described in detail in Smith, H.O. and Venter, J.C.
            (Making small insert libraries for whole genome shotgun
            sequencing projects. In Genome Sequencing: A Practical

```

```

Approach, eds. M. Vaudin and B. Barell, Oxford University
Press, 1999)."
BASE COUNT      203 a   122 c   170 g   138 t   1 others
ORIGIN

```

```

Query Match      9.3%; Score 33.4; DB 112; Length 634;
Best Local Similarity 58.6%; Pred. No. 6.9;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 250 ccaacagctgttgctccacgtaccactctctaaacacaggaagcactgagcagcg 309
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 348 CCAACAGTTTATCCCTTCAGGGACCACTTCTTACACCACAGCAGCAGGAGAATGG 407
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 310 caatcagtcaccacacagtcgtttgaggagagaagcgcgc 348
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 408 CTCTGTGATTCGCTTGAGACGGTGAAGGAGGTGAAGCC 446
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 10
LOCUS      BE430706      719 bp      mRNA      EST      24-JUL-2000
DEFINITION SUN007.C09F991221 ITEC SUN Wheat cDNA Library Triticum aestivum
            cDNA clone SUN007.C09, mRNA sequence.
ACCESSION  BE430706
VERSION     BE430706.1 GI:9428661
KEYWORDS   EST.
SOURCE     Bread wheat.
ORGANISM   Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
REFERENCE  1 (bases 1 to 719)
AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
            International Tritecae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
            Unpublished (2000)
            Contact: Shariflou M
            Plant Breeding Institute, University of Sydney
            107 Cobbitty Rd., Cobbitty NSW 2570 Australia
            Tel: 61 2 9351 8803
            Fax: 61 2 9351 8851
            Email: msharif@pop.usyd.edu.au
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.

```

```

TITLE     Expressed Sequence Tags for Species of the Triticeae
JOURNAL   Unpublished (2000)
COMMENT   Contact: Shariflou M
            Plant Breeding Institute, University of Sydney
            107 Cobbitty Rd., Cobbitty NSW 2570 Australia
            Tel: 61 2 9351 8803
            Fax: 61 2 9351 8851
            Email: msharif@pop.usyd.edu.au
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.

```

```

FEATURES             Location/Qualifiers
     source           1..719
                     /organism="Triticum aestivum"
                     /db_xref="taxon:4565"
                     /clone="SUN007.C09"
                     /clone_lib="ITEC SUN Wheat cDNA Library"
BASE COUNT      115 a   264 c   184 g   142 t   14 others
ORIGIN

```

```

Query Match      9.3%; Score 33.4; DB 35; Length 719;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 64; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```

```

QY 63 ctcaacttcgcgaactccgcagctgtctgatttcgacgtctattcgagctgtttctaa 122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 330 CTCACCTCCGCTTTTCGGCTTCTCCGACTCCGGCCGGGATTCACAGGTCTACGG 389
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 tgcattactgtctacataccacgggtagtcgctatgcccccgctcaagcgc 177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 390 CGAGCTATTCGACAAGGTATACGCCAGGAGTGCCTTCATGCCCGCCGATGGC 444
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 11



```

AA334604/C
LOCUS      AA334604      331 bp      mRNA
DEFINITION EST38844 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION  AA334604
VERSION     AA334604.1
KEYWORDS    EST.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
REFERENCE   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
AUTHORS     C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D., White
            O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodok,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palancques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
            Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE     96026280
COMMENT     Other ESTs: THC174272
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
            Location/Qualifiers
                1. 331
                /organism="Homo sapiens"
                /db_xref="ATCC (inhost):136288"
                /db_xref="taxon:9606"
                /clone_lib="Embryo, 9 week"
                /dev_stage="embryo, 9 wks"
                /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                XhoI"
BASE COUNT  89 a 72 c 81 g 89 t
ORIGIN
Query Match 9.2%; Score 33; DB 3; Length 331;
Best Local Similarity 51.0%; Pred. No. 8.2;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Qy 153 agtgcctatgccccgcgtcaagcgagcagcctaaacccggctgttaagcaagatgagca 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 AGTTCAGCAGACCCCTGTACAGCTGTTCCTCCGCCCACTGGGATCCTTAAGGAACAGGCTG 97
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 213 gaagcccaaacggaagcgctcaattggctgttaagcccaacagctgttggcgctccact 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 GAAGAGCCAGCGGCGCTCCCTCAATAATGCCTAAATAACCAACATATCTATGCCGACGCC 37
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 273 accattctcttaaaaaaacggaagcactggagcc 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36 ATCCCTCTCTCAATGCCATGGTCCANGGGAGTCC 4
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
LOCUS      AA859269/C
DEFINITION AA859269 489 bp mRNA EST 03-JUL-1999
            UI-R-E0-cj-g-01-0-UI-s1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-cj-g-01-0-UI-3' similar to dbj|AB002381|AB002381 Human mRNA
            for KIAA0383 gene, partial cds, mRNA sequence.
ACCESSION  AA859269
VERSION     AA859269.1
KEYWORDS    GI:4229623
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
AUTHORS     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 489)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948620.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
            Bonaldo, Ph.D. Clone distribution: clones will be available through
            Research Genetics This clone is also available through the
            I.M.A.G.E. Consortium at LNL (info.image.lnl.gov). IMAGE
            ID=I777539 The following repetitive elements were found in this
            cDNA sequence: 1-25, >AT-rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
            Location/Qualifiers
                1. 489
                /organism="Rattus norvegicus"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-E0-cj-g-01-0-UI"
                /clone_lib="UI-R-E0"
                /dev_stage="embryonic"
                /lab_host="DH10B (Life Technologies)"
                /note="Vector: pT73D-Pac (Pharmacia) with a modified
                polylinker; Site_1: NotI; Site_2: EcoRI; This library
                consists of a mixture of individually tagged normalized
                libraries constructed from 8, 12 and 18-day embryo. The
                tag is a string of 3-5 nucleotides present between the
                Not I site and the oligo-dT track which allows
                identification of the library of origin of a clone within
                the mixture."
BASE COUNT  87 a 108 c 138 g 156 t
ORIGIN
Query Match 9.2%; Score 33; DB 6; Length 489;
Best Local Similarity 48.2%; Pred. No. 8.8;
Matches 93; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 156 gccgtatgccccgcgtcaagcgagcagcctaaacccggctgttaagcaagatgagcagaa 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 GCTACATGAATACAGACACCCAGATACCCCTATGCAGATGCAGATGGCATGTGGCACC 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 216 gcccaaacggaagcgctcaattggcggttaagcccaacagctgttggcgctccactacc 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 AGCCATACGCCAGCAGCCCAATGCAGACCCACCCACAGTACATGATGTACAGGCC 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 276 acttcttaaaaaaacggaagcactggagcgagcgcaatcagtcctcccaacagctgttgg 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 216 CCGACACCGGTACATGAACACGGGATGTCCTCAACAGCTCTCAACGGGTCTTACA 157
QY 336 ggagaagccgcc 348
Db 156 TGAGAAGGTAGAC 144

RESULT 13
AW955769/c
LOCUS
DEFINITION EST 01-JUN-2000
ACCESSION EST367839 MAGE resequencences, MAGD Homo sapiens cDNA, mRNA sequence.
VERSION AW955769
KEYWORDS AW955769.1 GI:8145452
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 713)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gasparid,R., Gay,C., Holt
I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 83
Seq primer: Reverse.
FEATURES
source Location/Qualifiers
1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequencences, MAGD"
/notes="Vector: pBluescriptSKm"
BASE COUNT 196 a 133 c 167 g 217 t
ORIGIN

Query Match 9.2%; Score 33; DB 25; Length 713;
Best Local Similarity 51.0%; Pred. No. 9.4;
Matches 78; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 153 agtgcgctatgccccgcgtaagcgcgagcagcctaaaccgctgttaagcaagatgagca 212
Db 155 AGTTCAGCAGACCTTGTACAGCTGTTCCTCGGCCAACTGGGATCCTTAAGGAACAGCGTG 96
QY 213 gaagcccaaacgcgacgctcacattgggtgttaagcccaacagctgttgccgctccacgt 272
Db 95 GAAGAGCAGCGGCGCTCCCTCACAATAATGCTTAAATACCAACATATCTATGCCACAGCC 36
QY 273 accacttcctaaacaggaagcactggagcc 305
Db 35 ATCCCTCTCTCAATGCCATGTGTCAGGGAGTC 3

RESULT 14
CNS00KK2 1101 bp DNA GSS 03-JUN-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR17F22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL077673
VERSION AL077673.1 GI:4957249
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR17F22"
/notes="end : T7"
BASE COUNT 191 a 94 c 163 g 151 t 502 others
ORIGIN

Query Match 9.2%; Score 33; DB 121; Length 1101;
Best Local Similarity 8.8%; Pred. No. 10;
Matches 27; Conservative 144; Mismatches 136; Indels 0; Gaps 0;

QY 15 cagcgcgctgagagggtgtaagaaggttctgttgaggcaagcaactcaactccgc 74
Db 795 MGGMMMMVMMGMMGGMGKGHMMGKMGKTMRRKMYNMKMGMAKMMBMMAMMMMMK 854
QY 75 aactccgacgtgtctgcttcgacgtcttcgaagcttcttaactgcttacttct 134
Db 855 MMNKDKNNWCKTKOMNNKMYMMMMMMMAKMMHMKMMKMMKMMHMKMMHMKMMGTCMG 914
QY 135 cgtacactaccacccgggtagtcgctgtagcccccgcagcgcgagcagcctaaaccgc 194
Db 915 CKKKMAKTCBMMMAKMAKVGSMAKDKMDGMAGMAMMNCMCGCMGMMAGCKMKMMMMG 974
QY 195 tgttaagcaagatgagcagaagcccaaacgcgacgctcacattggctgtttaagccaac 254
Db 975 AGVAMGGTAGKMMKMGTMAMMAKMAKMAKMAKMAKMAKMAKMAKMAKMAKMAKMAKMA 1034
QY 255 agctgttgccgtccacgtaccacttcctaaacagcaagcactggagcagcgaatc 314
Db 1035 MKMKHVMHMMKMMMDHANNMKMACMMNMKMMAGKKKKMMVMMMMMKMMVNDGKMMMAK 1094
QY 315 agtccca 321
Db 1095 MADKMM 1101

RESULT 15
CNS00TNV/c
LOCUS
DEFINITION CNS00TNV 480 bp DNA GSS 28-JUN-1999
Arabidopsis thaliana genome survey sequence T7 end of BAC T401 of
TAMU library from strain Columbia of Arabidopsis thaliana, genomic
survey sequence.
ACCESSION AL089929
VERSION AL089929.1 GI:5291069
KEYWORDS GSS.
SOURCE thale cress.

```

```

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophyta;
Eukaryophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 480)
Salanoubat,M., Choisine,N., Artiguenave,F., Brottier,P., Wincker,P.,
Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
Unpublished
2 (bases 1 to 480)
Genoscope.
Direct Submission
Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
Location/Qualifiers
1..480
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="TAMU"
/clone="T401"
/note="end : T7"
140 a 100 c 103 g 137 t
BASE COUNT
ORIGIN

```

	Query Match	9.1%	Score 32.6;	DB 121;	Length 480;
	Best Local Similarity	55.9%;	Pred. No. 12;		
	Matches 62;	Conservative 0;	Mismatches 49;	Indels 0;	Gaps 0;
QY	52	gaggcaaggacctaacttcgcgaacttcgcagcgtctctgtatttcgacgtcgtattcgaa	111		
Db	322	GTGAAAAAGGTCGAAGGAGCGCGTCACACCGTCGTACCCCTGTCTATCAGAAATTC	263		
QY	112	gctgtttcttaatgcataacttctgtcgtacataccacacgggtagtgcggtat	162		
Db	262	GATGTTCTCTGATTCTCTCTGGTCTTTCTAGATCACAAGGTAGAGCCCAAT	212		

Search completed: December 18, 2000, 15:12:46  
Job time: 18733 sec